

Use of Experimental Methods in Workforce Evaluations

**Gary Burtless, Brookings Institution
David H. Greenberg, University of Maryland**

This report has been funded, either wholly or in part, with Federal funds from the U.S. Department of Labor, Employment and Training Administration under Contract Number AF-12985-000-03-30. The contents of this publication do not necessarily reflect the views or policies of the Department of Labor, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

Table of Contents

1.0	Introduction—When Is an Experiment Worthwhile?	1
1.1	Advantages of Randomization	2
1.2	Disadvantages and Threats to Experimental Validity.....	7
1.3	Implications for Experimentation	14
2.0	Past and Ongoing Experiments	15
2.1	Target Group Typology	17
2.2	Program Area Typology	20
3.0	Lessons from Past and Current Experiments	30
3.1	Feasibility, Cost-Effectiveness, and Time Costs	30
3.2	Have Experimental Results Influenced Workforce Policy?	33
3.3	Welfare-To-Work Experiments.....	33
4.0	Some Suggestions for Social Experiments.....	36
4.1	Higher Priority Experiments	36
4.2	Lower Priority Experiments	44
4.3	Postscript	47
	Endnotes	48
	Appendix A: Topologies and List of Experiments	52
	Appendix B: Summaries of Selected Social Experiments	72

1.0 Introduction—When Is an Experiment Worthwhile?

Randomized trials are an increasingly popular method for evaluating policy alternatives in a variety of fields. The growing popularity of experiments is due to a simple fact. Use of random assignment increases the likelihood that an evaluation will produce a valid estimate of the impact of a policy intervention.

This report considers the pros and cons of using randomized trials to improve the effectiveness of workforce policies. Experimentation can be used both to assess the effectiveness of current policies and to develop reliable information about the likely effects of new policy innovations. Of course, randomized trials are not always feasible, and even where feasible they do not necessarily provide reliable evidence about policy impacts. But a wide range of public policies are aimed at improving labor market outcomes at the individual level, and for this kind of program a randomized experiment has important advantages over other methods of policy evaluation. In particular, a randomized trial offers researchers a convincing benchmark for measuring program effectiveness. An experiment allows us to answer a question that is usually unanswerable in nonexperimental studies: How would people enrolled in the tested program have fared if they had not been offered services under the program?

The report is divided into four main sections. In the remainder of this section 1.0, we define randomized trials and describe their advantages and disadvantages in comparison with other techniques for learning about program effectiveness. We describe the circumstances in which experiments should be used and those in which they should be avoided. In section 2.0, we review previous use of social experiments in workforce evaluation. Our review provides guidance in thinking about what kinds of experiments should be conducted in the future. By examining the target groups and the types of programs that have previously been the subject of social experiments, we can also identify areas where major gaps remain. What target groups have been missed in past and current experiments assessing workforce interventions? What kinds of treatments have not been subject to experimental evaluation?

Section 3.0 considers lessons from past workforce experiments. We are particularly interested in learning about policy questions and the kinds of workforce interventions where experimentation has proved most fruitful. Have past experiments produced knowledge that is useful for evaluating and improving policy? Under what conditions has knowledge obtained in experiments been used to change policy? Did the results of experiments offer a reliable guide to impacts when tested policies were adopted?

In the last section 4.0, future social experiments that the Department of Labor (DOL) might sponsor are considered. We develop two lists of potential experiments. The experiments in the first list should have high priority for the future improvement of workforce policy. The proposed experiments in the second list, while worthwhile, have lower priority either because they face greater technical challenges or because they deal with less important research questions. Recommendations are based on our judgment of DOL's research priorities as well as on our assessment of the successes and failures of previous experiments. Both technical issues and political feasibility when developing recommendations have been taken into account, and

suggested experiments that are likely to be particularly difficult to implement are indicated. In discussion of high-priority experiments, how a randomized trial might be designed and the possible limitations of the research findings are described.

1.1 Advantages of Randomization

The critical element that distinguishes controlled experiments from all other research methods is the random assignment of meaningfully different treatments to the observational units of study. A randomized field trial (or social experiment) is simply a controlled experiment that takes place outside a laboratory setting, in the usual environment where social and economic interactions occur. In the simplest kind of experiment, a single treatment is assigned to a randomly selected group (the treatment group) and withheld from the remainder of the enrolled sample (the control or null-treatment group). Many social experiments have tested a variety of different treatments rather than only one. Some have not enrolled a pure control group at all. Instead, investigators have attempted to measure the differences in effect of two or more distinctive new treatments. The definition of an experiment can include tests of innovative new policies as well as studies that are intended to measure the impact of a current policy relative to a null treatment.

In the absence of information from social experiments, analysts rely on four main kinds of information to learn about the effectiveness of particular programs or interventions. One source of information is data on the relationship between nationwide or communitywide aggregates, such as average educational attainment, on the one hand, and median wage earnings on the other. The effect of treatment differences is inferred by measuring the outcome differences over time or across communities. However, aggregate statistics are inappropriate for analyzing many kinds of detailed relationships, such as the effectiveness of different instructional methods in workforce training programs. A second source of information is management data collected in the administration of existing programs. Information from an existing program almost never provides any data about what the participants' experiences would have been if they had been enrolled in a different program or in no program at all, however. This greatly limits the usefulness of such information for inferring the impact of program services on participants' behavior or job market success.

A third source of information is new survey data. Such data are more costly to obtain than existing programmatic data, but they provide information about the experiences of nonparticipants as well as participants in a program, and thus offer some evidence about behavior or outcomes in the absence of treatment. A fourth source of information is data generated by special demonstration programs. Like experiments, demonstrations involve the special provision of a possibly novel treatment, data collection about behavioral outcomes, and analysis of treatment effects. Unlike experiments, demonstrations do not necessarily involve random assignment. (When research projects are called "demonstrations" but involve random assignment, we classify them as "experiments.") What all experiments have in common is that the tested treatments are randomly assigned to observational units—that is, to individuals, to companies, to job service offices, or to entire communities.

The usual focus of analysis in an experiment is on one or more measurable outcomes that can be observed within the experimental sample at the individual level. Outcomes of interest to the DOL include a sample member's employment status, duration of employment, hourly or weekly earnings, collection of public transfer benefits, duration of unemployment, and attainment of some educational or training credential. The goal of the analysis is to assess the impact of a policy intervention on one or more of these outcomes. Random assignment can also be used to evaluate administrative processes or management costs, but it is rarely worthwhile to use an analysis technique as costly as experimentation to obtain this kind of information. For example, a randomized trial could be designed to test new cost-saving methods of taking unemployment insurance applications or recruiting training candidates for workforce programs. It is usually feasible to assess these kinds of management innovations reliably with evaluation methods much less expensive than randomized trials, however. Less expensive methods are also available to analyze the influence of DOL policies on employment, earnings, and job market credentials, but the alternative research methods are usually far less reliable than experimentation in determining the behavioral effects of a policy.

Why Is An Experiment Worthwhile?

The advantages of controlled experiments over other methods of analysis are easy to describe. Because experimental subjects are randomly assigned to alternative treatments, the effects of the treatments on behavior can be measured with high reliability. The assignment procedure assures us of the direction of causality between treatment and outcome. Differences in average outcomes among the several treatment groups are caused by differences in treatment; differences in average outcome have not caused the observed differences in treatment. Causality is not so easy to establish in nonexperimental data. A simple comparison between participants and nonparticipants in a government training program might reveal, for example, that program participants subsequently earn lower wages than eligible workers who never enrolled in the program. Does this indicate the program was unsuccessful? Or does it show instead that the program was very attractive to people who expect to earn very low wages in the future? The poor earnings prospects of some workers might cause them to enroll. Alternatively, inappropriate or low quality training might have caused participating workers to earn below-average wages. If eligible workers are assigned at random to the training program, the direction of causality would be known with certainty.

Random assignment also removes any systematic correlation between treatment status, on the one hand, and observed and unobserved participant characteristics on the other. Estimated treatment effects are therefore free from the selection bias that potentially taints all estimates based on nonexperimental sources of information (see below). In a carefully designed and well-administered experiment, there is usually a persuasive case that the experimental data have produced an internally valid estimate of average treatment effect.¹

Another advantage of experiments is that they permit analysts to measure—and policymakers to observe—the effects of environmental changes and new kinds of treatment that have not previously been observed. In many cases, the naturally occurring variation in policies is too small to allow analysts to infer reliably the effects of promising new policies. Some people believe, for example, that job search programs would achieve better results if managers in these

programs received money bonuses linked to the job-finding success of the unemployed workers they serve. If a bonus system of this kind has never been tried, it is impossible to predict how managers would respond to a particular bonus scheme or how job seekers would fare under such a system. Of course, new policies can be tested in nonexperimental demonstration projects, too. But in comparison with most sources of nonexperimental information, experiments permit analysts to learn about the effects of a much wider range of policies and environmental changes.

The Canadian Government recently examined the feasibility of offering public assistance recipients a substantial incentive to take full-time jobs. The program, known as the Self-Sufficiency Project, offered to pay public aid recipients one-half the difference between their actual wages and roughly \$44,400 (U.S. \$37,000) per year.² Canadian aid recipients could only receive the wage supplement if they worked in a full-time job and entered the job within one year of being offered the supplement. Because Canada had never previously offered a work incentive even remotely this generous, the government would have found it difficult to predict reliably the consequences of the incentive using information available when the Self-Sufficiency Project was launched. Wisely, Canadian officials decided to test the supplement in a small-scale randomized trial.

Straightforward Interpretation of Results

The simplicity of experiments offers notable advantages in making results convincing to other analysts and understandable to policy makers. A carefully conducted experiment permits analysts to describe findings in extremely straightforward language: “Relative to unemployed workers in the control group, unemployed workers eligible for employment bonuses found jobs faster and reduced the average duration of their unemployment spells by three weeks.” This kind of simplicity in describing results is seldom possible in nonexperimental studies, where research findings are subject to a lengthy and often confusing list of qualifications.

In recent years, the last advantage of experiments has turned out to be particularly important for experiments that test practical policy alternatives in adult training and welfare policy. Because policymakers can easily grasp the findings and significance of a simple experiment, they concentrate on the implications of the results for changing public policy. They do not become entangled in a protracted and often inconclusive debate about whether the findings of a particular study are statistically valid. Voters and politicians are more likely to act on results they find convincing.

Two experiments provide vivid illustrations of the persuasive power of randomized trials. In the late 1980s, DOL followed the advice of social scientists and conducted an experimental evaluation of the Job Training Partnership Act (JTPA). The JTPA experiment produced convincing evidence that many DOL-funded training programs for 16- to 22-year-olds were ineffective in boosting participants’ earnings.³ This evidence was found so persuasive that the Clinton administration proposed drastic reductions in spending on the young target population, suggesting a reallocation of funds to groups where training was shown to be effective. Congress accepted the administration’s reallocation with little objection, because the experimental findings left little reason to think the programs targeted on youth were achieving the desired effect. Critics of the administration’s proposal could not point to methodological doubts about the results in

arguing for continued funding of youth training under the JTPA program. Even when analysts disagree about the policy significance of findings from a randomized trial, it is usually impossible to reject the internal validity of the estimated treatment impact uncovered in the experiment. Economists James Heckman and Alan Krueger disagree about the policy implications of the JTPA experimental findings, but both acknowledge that the experiment showed no gain in earnings among the young people enrolled in the JTPA treatment group.⁴

A more hopeful illustration of the power of randomized trials is the recent experience of Minnesota with its Family Investment Program for public assistance recipients. Like the Canadian Self-Sufficiency Project, the Minnesota program tested the effects of providing more generous incentives for aid recipients to move into full-time jobs. Under the previous Minnesota welfare program, parents who went to work typically lost most of their cash assistance benefits because their monthly grants were reduced nearly one dollar for every dollar they earned in a job. The Family Investment Program allowed recipients to keep most of their grants for an extended period, even after they found jobs that paid modestly good wages. When this incentive was combined with required participation in job-search and other employment-related activities, participants' employment rates increased by one-third and their earnings increased by one-quarter.⁵ If these results were observed in a nonexperimental evaluation, they would have elicited little comment and might not even have been believed. Because they were observed in a randomized trial, news reporters and political observers accepted the findings at face value and focused on their implications for welfare policy. The findings were easy to believe and even easier to understand. They did not become the focus of a protracted debate about the believability of the research methodology that produced them. Minnesota has expanded the program tested in the experiment to cover its entire welfare caseload, and other states have followed Minnesota's example.

Of course, an experiment does not completely eliminate our uncertainty about the answer to a well-posed question concerning individual behavior or policy effectiveness. But it can dramatically reduce our uncertainty. Equally important, the small number of qualifications to experimental findings can be explained in language that is accessible to people without formal training in statistics or the social sciences. This is a crucial reason for the broad political acceptance of findings from recent labor market experiments.

Sample Selection

Sample selection is one of the most common sources of bias in statistical studies that focus on individual behavior. Nonexperimental studies of education and training, for example, usually rely on observations of naturally occurring variation in treatment doses to generate estimates of the effects of training. Analysts typically compare measured outcomes in test scores, employment, or earnings for participants in an educational or training program and for a comparison group of similar people who did not benefit from the educational or training intervention. For example, the value of a college degree is often calculated by comparing the earnings of college graduates with the earnings of people born in the same year who graduated from high school but did not attend college. Even if the analysis adequately controls for the effects of all measurable characteristics of people in the sample, it is still possible that average outcomes are affected by systematic differences in the unmeasured characteristics of people in

the treatment and comparison groups. In the simplest kind of adult training program, for example, people in the sample are exposed to just two potential doses of treatment, say, $T = 1$ for people who enroll in the training program, and $T = 0$ for people who do not enroll. Program participation represents the sample member's decision to choose one treatment dose or the other. Obviously, this decision may be affected by the person's unobserved tastes or other characteristics, which in turn may also affect the person's later employment and earnings. Since these factors are unobserved, their effects cannot be reliably determined. The exact amount of bias in the nonexperimental estimate thus remains uncertain.

Selection bias is a practical estimation problem in most nonexperimental policy studies. Naive analysts sometimes ignore the problem, implicitly assuming that unmeasured differences between program participants and nonparticipants do not exist or do not matter. While one or both of these assumptions might be true, the case for believing either of them is usually weak. Critics of early training evaluations sometimes pointed out, for example, that people who voluntarily enroll in employment training for the disadvantaged might be more ambitious than other disadvantaged workers. Unfortunately, ambition is an unmeasurable personal trait. If personal ambition is correlated with a person's later labor market success, it is unclear what percentage of the average earnings advantage of trainees is due to extra training and what percentage is due to the greater average ambition of people who apply for and receive training.

Selection bias can go in the opposite direction. Disadvantaged students who are unlikely to fare well in the job market may be disproportionately likely to enroll in after-school or summer programs that are aimed at increasing students' employability. If their voluntary decision to enroll in the after-school program is based on a valid, but unobserved, assessment of their earning potential, we would expect that their wages in the absence of the after-school program would be lower than those of students with identical observable characteristics who decide against enrolling in the program.

Our uncertainty about the presence, direction, and potential size of selection bias makes it difficult for analysts to agree on the reliability of estimates drawn from nonexperimental studies. The estimates may be suggestive, and they may even be helpful when estimates based on many competing research strategies all point in the same direction. But if statisticians obtain widely different estimates or if the available estimates are the subject of strong methodological criticism, policy makers will be left uncertain about the genuine effectiveness of the program. In the case of adult training programs, Orley Ashenfelter and David Card (1985) and Burt Barnow (1987) showed that this kind of uncertainty was not just a theoretical possibility. Both sets of reviewers found that the nonexperimental literature contained an unhelpfully wide range of estimated impacts of the Comprehensive Employment and Training Act (CETA) on earnings.⁶ The range of estimates reported in the nonexperimental studies was too wide to permit policy makers to decide whether CETA-sponsored training was cost effective, particularly in the case of men who received services under the program. It was not clear for some groups whether the impact of CETA training was beneficial at all. As noted above, the findings from the later experimental study of Job Training Partnership programs were not subject to the same kinds of criticisms. They were produced in a randomized trial rather than with nonexperimental research methods. Economists and other researchers have broadly accepted the findings of the Job Training Partnership experiment, and policymakers have acted upon them.

1.2 Disadvantages and Threats to Experimental Validity

Experimental evaluations, like other research methods, can have important shortcomings. Under some circumstances, these disadvantages may be so serious that a randomized trial offers little or no advantage compared with nonexperimental methods. In considering the list of disadvantages associated with randomization, however, it is worth bearing in mind that many nonexperimental methods are subject to identical or very similar limitations. The relevant question for a particular policy question is: What are the costs and benefits of an experiment compared with the feasible research alternatives?

Cost Considerations

Experiments have three kinds of costs that, under certain circumstances, make them more expensive than most nonexperimental research on the same topic. This is especially likely if the nonexperimental research is an evaluation of an existing policy. Not surprisingly, running a controlled experiment consumes more real resources than a nonexperimental study that uses existing data sources. An experiment is also costly in terms of time. Several years usually elapse between the time an experiment is conceived or designed and the release of its final report. If policy decisions about a particular policy question cannot be deferred, the usefulness of an experiment may be questionable.

In addition, experiments often involve significant political costs. It is more difficult to develop, implement, and administer a new treatment than it is simply to analyze information about past behavior or collect and analyze new information about existing policies. Voters and policy makers are rightly concerned about possible ethical issues raised by experiments. As a result, it is usually easier to persuade officials to appropriate small budgets for pure research or medium-sized budgets for a new survey than it is to convince them that some people should be systematically denied a potentially beneficial intervention in a costly new study.

These disadvantages of experiments are real, but it is easy to exaggerate them. Some forms of nonexperimental research suffer from identical or very similar disadvantages. A demonstration study, which lacks a randomly selected control group, can easily cost as much money as a social experiment that tests the same treatment. A nonexperimental demonstration will almost certainly take as much time to complete as an experiment. If Canada had elected to study the work incentive in its Self-Sufficiency Project using a demonstration instead of a randomized trial, it would have saved little money and no time. If a new survey must be fielded to obtain needed information, the extra time and money required for an experiment may be relatively modest.

The Black Box Issue

A criticism that is made of some social experiments is that the evaluated treatment consists of a combination of components—for example, job search, training, and financial incentives—and it is not clear how each component contributed to any measured effects. This is often called the “black box problem.” This is also a shortcoming of many nonexperimental evaluations, although econometric methods are sometimes used to try to distinguish among the effects of different treatments. Once the overall effects of a program are measured, an

experimental evaluation can also use these nonexperimental techniques to attempt to obtain additional information about the contributions of various program components.

More importantly, however, experiments provide a much better opportunity to enter the black box than other evaluation methods because it is sometimes possible to randomly assign individuals to different treatment groups that offer different components. For example, in both the Canadian Self-Sufficiency Project and the Minnesota Family Investment Program, which were discussed earlier, some individuals were randomly assigned to a group that offered only financial incentives, while others were assigned to a group that provided services as well as financial incentives (a third set of individuals were assigned to a control group). In the evaluation of both of these programs, the combination of financial incentives and services performed better than the financial incentives alone.

Ethical Issues Connected With Human Experimentation

Many policymakers and voters are concerned with the ethical issues raised by experiments involving human subjects, especially when the experimental treatment (or the denial of treatment) has the capacity to inflict serious harm. If the tested treatment is perceived to be beneficial, program administrators may find it hard to deny the treatment to a randomly selected group of people enrolled in a study. Except among philosophers and research scientists, random assignment is often thought to be an unethical and even inhumane way to ration scarce resources. If the treatment to be tested is viewed as harmful, it may be difficult to recruit program managers to run the project and impossible to recruit voluntary participants. Thus, few such experiments are conducted. It may not be ethical to mount such an experiment in any event. Readers should recall, however, that similar ethical issues arise in studies of new medicines and medical procedures, where the stakes for experimental participants are usually far greater than they are in employment and training experiments. Yet randomized field trials have been common in medicine for much longer than has been the case in social policy. In fact, randomized trials are frequently required as a matter of law or regulation in order to demonstrate that a new medical treatment is safe and effective.

Good experimental design can reduce ethical objections to random assignment.⁷ At a minimum, participants in experimental studies should be fully informed of the possible risks of participation. Under some circumstances, people who are offered potentially injurious treatments or denied beneficial services can be compensated for the risks they face. The risks of participation are usually unclear, however, because it is uncertain whether the tested treatment will be beneficial or harmful.

Of course, uncertainty about the direction and size of the treatment effect is the main reason that an experiment is worthwhile. A successful experiment substantially reduces our uncertainty about the size and direction of the treatment effect, including potentially adverse impacts. The ethical argument in favor of experimentation is that it is preferable to inflict possible harm on a small scale in an experimental study rather than unwittingly to inflict harm on a much larger scale as a result of misguided public policy.

Threats to Experimental Validity

Experimental studies face two kinds of threats to drawing valid inferences from their statistical results. The first kind of threat involves making erroneous inferences about treatment impacts within the sample actually enrolled in the experiment. The second involves making erroneous inferences about the applicability of the experimental findings to the general population represented by the enrolled sample. We refer to the first kind of problem as “within-sample errors of inference” and to the second as “out-of-sample errors of inference.”

Although randomized trials are frequently criticized for making within-sample errors of inference, the errors that critics point to almost always represent potential problems for all social science research methods.⁸ One common problem, for example, is sample attrition in the experimental sample. Random assignment to treatment should guarantee that any observed differences in outcomes between treatment groups are attributable to differences in treatment. For this to hold, outcomes must be measured for all, or a random subset of, enrolled sample members who are assigned to each treatment group. In practice, outcomes are usually measured with information obtained from in-person interviews or administrative records. In the case of interviews, some participants may refuse to answer a survey and others may not be located by interviewers. Administrative records often provide information for all or nearly all of the people enrolled, but such records can be incomplete, too. Public assistance records, for example, may contain information about people who remain on Aid to Families with Dependent Children (AFDC), but they may lack information about people who enroll in the Supplemental Security Income (SSI) program after leaving AFDC. Unemployment insurance wage records contain information about covered wages earned in a particular state, but they do not contain information about earnings not covered by the unemployment insurance system or earned in another state. These limitations in the interview data or administrative records mean that analysts may not have outcome information for all or a random subset of the people enrolled in the experiment. To the extent that the sample loss is nonrandom with respect to the error term in the outcome-prediction equation, the resulting analysis sample will suffer from the same type of selection bias that can occur in nonexperimental studies.

Even if the sample loss is random with respect to the error term, shortcomings in the data may prevent analysts from measuring all of the outcomes that are relevant for assessing the value of a tested treatment. Consider those studies that rely on unemployment insurance earnings records to infer the impact of a treatment on enrollees’ earnings. More than 90 percent of money labor earnings in the United States is reflected in the earnings reports contained in these administrative records. Even if it is the case that same percentage of labor earnings is reported for members of both the treatment and control groups, some fraction of earnings (and of the earnings change due to the treatment) will be omitted from the administrative records. In the case of an experimental site where many workers earn wages in a neighboring state, the percentage of omitted earnings can be large. When many enrollees earn wages in uncovered employment, the fraction of missing wages can be even higher. If analysts do not have accurate information about the outcomes of interest, they cannot accurately calculate the impact of the tested treatment. Consequently, benefit-cost studies based on the experimental estimates will be unreliable.

While nonrandom sample loss and errors in measuring outcomes represent serious causes for concern, the problems are common to all research involving information obtained from interviews or administrative records. They are not unique to randomized trials. All interview data

suffer from the fact that response rates are less than 100 percent, and sometimes much less. If the survey design calls for repeated interviews of the same respondents, response rates are often far below 100 percent. The limitations of administrative records and surveys are the same whether they are used in experimental or nonexperimental studies. The problems of nonrandom attrition and errors of measurement receive more attention in experiments than in nonexperimental studies because they represent the main threats to the internal validity of an experimental finding. In nonexperimental research, by contrast, these problems usually seem relatively minor compared with other, far more serious threats to the validity of treatment-effect estimates.

The more serious problems of experimental evidence involve drawing accurate inferences about the applicability of the findings for a broader population than the one that was enrolled in the study. The goal of a good experiment is to inform us of the likely consequences of adopting the tested policy if it is introduced on a statewide or nationwide basis. Many past experiments, including a majority of those listed in the appendix A tables, tested new programmatic interventions in small samples. As a result, some effects that would occur in a program that was rolled out on a large scale may be missed. For example, if the services provided by an ongoing program are perceived as beneficial, then some individuals who are initially ineligible to participate may adopt behaviours needed to qualify (an “entry” effect). On the other hand, some individuals who might otherwise have initiated a claim for benefits may decide not to do so if they will be required to meet work or training requirements (a “deterrent” effect). Such individuals are typically not included in research sample used in an evaluation, regardless of whether it is experimental or nonexperimental.⁹ Thus, entry and deterrent effects are usually not accounted for in impact estimates.

A different problem, but one again related to the size of the demonstration program that is being evaluated, occurs because some of the tested interventions would have been much more difficult to put into effect and administer if they had been implemented on a city-wide, state-wide, or national scale. The administrative burden of implementing a particular training or employment strategy on a nationwide scale can change the nature of the treatment received by program participants, possibly reducing the average impact of the treatment on participant outcomes. This risk seems much greater when the experimental treatment is designed and administered by people who strongly favor the adoption of the tested treatment. Their care and energy in providing the treatment may be much greater than what could be expected if administration of the treatment were placed in the hands of less interested civil servants or private-sector workers. If the tested intervention were to be implemented on a nationwide basis, however, it is likely that it would be administered by ordinary civil servants and private-sector workers.

A handful of past experiments have not tested new policies but have instead tried to measure the impact of an existing policy. For example, both the JTPA and Job Corps programs have been evaluated using random assignment methods. Applicants who were eligible for these programs were randomly assigned to treatment and control groups. People assigned to the treatment group obtained the services normally provided to program participants; people in the control group were told they could not receive program services for a specified period of time, such as 24 months. The impact of the programs on enrolled applicants was estimated by comparing outcomes in the treatment and control groups. In this kind of research design, there is

little doubt that analysts are measuring the impact of a treatment as it is typically administered by ordinary program administrators.¹⁰

One risk of experimentation in this kind of study is that program administrators will be forced to change some of their usual procedures in order to accommodate the need for randomization. This may affect enrollees' perception or experience of the treatment offered. This kind of risk can be minimized by reducing the portion of the applicant population that is assigned to the control group. Obviously, a much greater burden is placed on administrators if they are asked to assign 50 percent of program applicants to the control group rather than only two percent or five percent of people enrolled. If a large percentage of applicants are assigned to control status, administrators may have to recruit many additional applicants in order to run their programs at the historical level. If they fail to find extra recruits, their programs may run with a higher ratio of resources to enrollees than is the norm. If they succeed in finding extra recruits, the population served by the program may differ from what it would be in the absence of randomization. These considerations imply that a more reliable estimate of program effect can be obtained in an experiment that is conducted in many sites and assigns only a small percentage of applicants to the control group. This design assures the most reliable estimate of the nationwide impact of the program, produces the least disruption in the normal operation of the program, and yields the smallest change in the characteristics of people who receive program services. However, such an experiment may place more burdens on the evaluators to maintain the integrity of the random assignment design (a topic that is discussed below) than an experiment that is limited to only one or a few sites.

All experiments are temporary, yet the programs they test are intended to be permanent if they are eventually implemented as state or national policy. This is not a problem for randomized trials that examine the effects of an ongoing program compared with a null alternative. In other cases, the difference between a temporary experimental program and a permanent program that implements the same treatment can be important in affecting our interpretation of the experiment's findings. In the case of an experiment that tests innovative financial incentives, for example, the distinction between a temporary and a permanent incentive is often critical. A program that offers financial incentives for people to enroll in schools or obtain jobs may have very different effects on behavior if the incentives are known to be temporary compared with the situation in which they are believed to be permanent. A person who is enrolled in a three-year experiment that pays for college tuition, room, and board may be more likely to take quick advantage of the incentive compared with a person who knows that the scholarship offer will remain in effect over the indefinite future. An unemployed worker who is offered a wage subsidy equal to \$4.00 per hour of paid employment may be more likely to immediately find and accept a job if the subsidy ends in three years than if the subsidy will remain available over then next three decades. In both examples, the temporary experiment dramatically changes the relative reward of acting sooner rather than later, whereas a permanent program does not change this relative reward.

Where this distinction between temporary and permanent incentives is important, experimental designs can be crafted to shed light on its importance. In the Seattle-Denver Income Maintenance Experiments (SIME/DIME), for example, the designers randomly assigned most of the sample to an experimental treatment that lasted three years. They also assigned a

smaller percentage of enrollees to identical treatments that lasted five years and a very small percentage of enrollees to treatments expected to last twenty years. The importance of short-duration bias can be inferred by comparing the treatment responses of people enrolled in the shorter duration and longer duration treatments.¹¹ In many cases, it is unclear whether the brief duration of an experiment really represents a problem. For treatments like job search assistance, lump-sum employment bonuses, and job training, the participant receives the entire experimental treatment within a few weeks or months. The duration of the experiment is essentially irrelevant to the participant's response to the tested treatment, because the entire response to the offered treatment is observable within a short period of time (although the effects of the treatment may, of course, remain for some time). Moreover, there is little reason to believe the response to the treatment will differ depending on whether the policy was permanent or only temporary.

Even in cases where short-duration bias poses a problem for inference, it can represent an important advantage for measuring crucial outcomes of interest. We noted earlier that time-limited incentives to encourage people to invest in extra education or training can elicit a faster response than would occur if the same financial incentive were believed to be permanent. In some cases, however, the size of this response may have only secondary interest. Policymakers' main interest may be in determining the effects of extra schooling and training on labor market outcomes. If a temporary incentive induces a faster and bigger investment in education or training, the human capital difference between the treatment and control groups will be magnified. Researchers will have a better chance of determining whether extra education and training leads to improvements in labor market outcomes for the enrolled population. The temporary nature of the incentive actually offers an advantage for measuring the effect that interests policymakers.

We believe a more serious problem for drawing valid inferences from experiments arises from the longer-term, general equilibrium response to a tested treatment. A small-scale experiment that tests an innovative policy cannot duplicate the informational environment that would develop if the same policy were implemented on a communitywide or nationwide basis. In an experiment, people who are enrolled in the treatment group must depend on the experiment's administrators to learn about details of the treatment. They cannot learn about advantages and disadvantages of different responses to the treatment from reading newspapers, observing the experiences of neighbors, or discussing advantageous strategies with friends. In a communitywide or nationwide program, these sources of information would eventually become available, and participants' improved knowledge about the treatment is likely to affect their behavioral response.

Changes in the informational environment may not only affect people's knowledge about a treatment and profitable responses to it, they may also influence the preferences of people who are exposed to the treatment. For example, a 50-percent subsidy for college costs may have a small impact on college attendance when almost none of your neighbors attend college but a much bigger impact when colleges enroll one-third or more of local high school graduates. College may seem like a more desirable goal when college attendance is common than when it is uncommon. If a permanent, nationwide program boosts college attendance in disadvantaged neighborhoods it may affect the perceived value of a college education in those neighborhoods. This long-term influence of the scholarship program on social norms is not observable in a

small-scale, short-term experiment. Readers should recognize, however, that nonexperimental research methods have almost no advantages over experiments in uncovering policy-induced changes in preferences. Standard statistical methods cannot reliably determine how much long-run change in behavior is due to economic agents' response to changed incentives and how much is due to changes in preferences induced by shifts in community norms.¹²

Small-scale experiments are designed to measure the direct effects of the offered treatment on the enrolled sample. They do not necessarily measure the full, general equilibrium response to a treatment that is permanently implemented on a communitywide or nationwide scale. For many kinds of treatment, the general equilibrium response is determined by reactions of many people or companies in addition to those represented in the experimental sample. Consider a reemployment bonus that provides Unemployment Insurance (UI) claimants with a large cash subsidy (say, \$5,000) if they find a job and leave the UI rolls within six weeks of first claiming a benefit. This kind of bonus strengthens the incentive for unemployed workers to accept a job during the first few weeks after claiming benefits. A small-scale experiment can measure the response of a handful of unemployed workers to this incentive. If the policy were adopted on a nationwide basis, every worker who claims UI would have an equally powerful incentive to quickly find a job during the first few weeks of unemployment. Stronger competition for job openings might reduce each claimant's job finding success compared with the success observed in a small-scale experiment.¹³

A nationwide program may also have effects on employers that are undetectable in a small-scale experiment. If only a handful of jobseekers are covered by the generous reemployment bonus, employers would be unlikely to encounter a covered jobseeker when considering applications to fill a vacant position. In contrast, when every new UI claimant is offered a reemployment bonus, a large fraction of job applicants will be covered by such a bonus. Employers might offer lower wages or less generous fringe benefits in response to the wide availability of job applicants who have a powerful incentive to accept immediate employment. This reaction of employers to the reemployment bonus will be missed in a small-scale experiment.

In considering whether an experiment is worthwhile, it is important to think carefully about the general equilibrium impact that the tested policy might cause if it were a permanent feature of the local or national environment.¹⁴ An experimental test of the policy may have little value if the response that it can accurately measure is likely to be a relatively minor part of the full general equilibrium response to a permanent policy. In the example we have just described, we think this is unlikely to be true. First, only a small percentage of unemployed workers are in the first 6 weeks of a covered unemployment spell, so the bonus we have described would directly affect the behavior of only a small fraction of active jobseekers. This means that queuing bias probably will not have an overwhelming effect on the response uncovered in a small-scale study. Second, economists and policymakers do not have enough evidence to predict whether a \$5,000 bonus would induce a large or a small response in job-finding success, even in the context of a small-scale study. Past reemployment bonus experiments have tested much less generous bonus formulas. Thus, it may be worthwhile learning whether generous bonuses could dramatically reduce the incidence of long unemployment spells.

To be sure, the experiment we describe cannot shed any light on employer responses to a generous reemployment bonus. Researchers will not learn whether wage offers would decline as a result of widespread availability of the bonus. Analysts would have to use nonexperimental evidence combined with the experimental findings to predict the general equilibrium effects of a generous bonus.¹⁵ Nonetheless, the experiment would provide much more reliable information about unemployed workers' responses to the incentive. In some cases, this kind of information is critical. For example, if a reemployment bonus experiment revealed that very few additional jobless workers obtained employment in the first six weeks of their job search, the employer response to a full-scale national program would have little interest. Since the bonuses failed to speed up reemployment in a small-scale study, it seems very unlikely they would have much influence on job finding in a national program. In this case, the employer response to subsidies is irrelevant, and the experiment has been successful in revealing the ineffectiveness of the tested policy.

1.3 Implications for Experimentation

Our discussion suggests a few main lessons for initiating and designing workforce program evaluations using experimental methods:

- Policymakers, voters, and researchers must be uncertain about the likely impact of the tested policy on an outcome of central interest. If the available research literature already provides persuasive evidence about the size and direction of this critical impact, resources should be devoted to examining some other policy question.
- The impact in question must be important in weighing the benefits and costs of the policy. It must be an impact where better knowledge of the size of the effect can decisively affect policymakers' and voters' assessment of the policy. It makes little sense to undertake an experiment that permits analysts to measure the impact of a policy on skill acquisition if voters and policymakers are instead concerned about its impact on employment and earnings. Voters and policymakers may have little interest in learning how many extra people acquire a General Educational Development (GED) diploma as a result of some new policy. If their interest focuses on the program's impact on employment and earnings, then it makes sense to design an experiment that can detect such an effect.
- Policymakers and researchers must be interested in outcomes for which tolerably accurate measurement is possible. Reliable measurement must be possible not only in theory, but also in practice. If a good benefit-cost analysis of the tested treatment requires accurate information about employment, wages, fringe benefits, the duration of unemployment spells, government transfer payments, or the acquisition of educational and skill credentials, the research design must identify data sources and procedures that will provide the needed information.
- The experimenters must identify administrators and management procedures that will preserve the integrity of the randomly assigned treatment and control groups. That is, they must specify practical methods for randomly assigning enrollees into treatment

and control groups and procedures for maintaining the distinctive treatments of the different groups. Members of the treatment group must be offered and then consistently provided the experimental treatment; members of the control group must be denied this treatment and consistently provided with the intended null treatment.

- Experimenters should be confident that they can provide enrollees in the treatment group with clear and reasonably complete information about the treatment they are offered. Moreover, they should have a plausible expectation that enrollees will understand the consequences that follow from different kinds of behavior. For example, unemployed workers enrolled in the reemployment experiment described earlier should have a clear understanding that they are eligible for a bonus if they find a job in the sixth week after claiming UI but are ineligible for a bonus if they find a job in the seventh week.
- Policymakers should be cautious about launching an experiment when theory or previous experience suggests that the general equilibrium response to the tested treatment will overwhelm or fully offset the likely response in a small-scale sample. It may not make sense to test an hourly wage subsidy for a large class of workers if available evidence suggests that employers will largely or fully offset the government subsidy by reducing the gross wage offered to that class of workers. Even though an experiment can provide an internally valid estimate of the impact of this kind of policy on enrolled workers, it is doubtful whether the findings can be validly extrapolated to a wider population.

We are confident that many workforce program experiments are feasible and worthwhile under these criteria. At a minimum, experiments aimed at assessing the value of ongoing programs meet all the requirements suggested in the above list.

2.0 Past and Ongoing Experiments

In this section, we review previous experience of social experimentation in workforce program evaluation. This kind of review can provide useful guidance in determining what sort of experiments should be conducted in the future. In particular, by examining the target groups and types of programs that have previously been and not been the subjects of social experiments, we can identify the areas where additional experimentation may not be very productive and where important gaps remain. There may, of course, be good reasons for the gaps. As discussed in the previous section, there are circumstances in which experiments do not provide very useful information to policymakers. For example, there may be technical or political reasons why it is difficult to operate certain programs under random assignment and, as a consequence, these programs have not been previously evaluated by social experiments. However, if this is not the case, future social experiments can potentially fill the gaps.

It is also possible that although certain target groups or program areas have been the subject of previous experiments, these experiments have not been technically sound or their results have been largely ignored by policymakers. A review of such experiments may suggest changes in design that might result in more successful experiments.

To examine previous social experiments, we have extracted information on experiments from the third edition of *The Digest of Social Experiments* (Greenberg and Shroder 2004). The *Digest* contains two to three page summaries of all the known randomized social experiments that were completed by the end of 2002 (240 in all) or were ongoing (i.e., planned, in the field, or in the final research phase) at that time. Summaries of 16 selected experiments that have been extracted from the *Digest* appear in appendix B of this report. Although not all the experiments reviewed were funded by the Department of Labor, we limited our review to the subset of experiments in the *Digest* that appear pertinent to the interests of DOL and, consequently, meet the following criteria:

- The evaluated programs provided services or incentives that were intended to improve labor market outcomes (for example, increase earnings or employment or reduce the length of unemployment).
- The experimental treatment was targeted at individuals of working age.
- The experiments were conducted in the United States.

A total of 150 social experiments met these criteria. These experiments are listed in appendix A of this report and are categorized on the basis of two typologies. The first of these typologies provides a set of target group categories and the second a set of program categories. These taxonomies are listed in table 1, which is found at the end of this section 2.0. More importantly, table 1 also provides a count of the number of experiments that fall into each category.

Although it was readily apparent how most of the experiments in our sample should be classified in terms of the typologies, there were a number of instances in which it was necessary to exercise our best judgment. This was particularly true of the program categories. Nonetheless, we are confident that table 1 provides a reasonably accurate picture of which categories have frequently been subject to social experimentation and which have not.

In categorizing experiments according to the target group category, assignment was on the basis of the group or groups at which the tested policy was specifically aimed. On the one hand, for example, a program aimed at unemployed adults was assigned only to the unemployed adult category, even though some of these persons are also on welfare recipients, disabled, or substance abusers. On the other hand, a program that was more narrowly targeted on welfare recipients or young people in welfare families was assigned only to the welfare recipient category. However, a few experiments tested policies that were specifically aimed at several of the target groups listed in table 1. For example, the training program run under the National Job Training Partnership Act was specifically targeted at both unemployed adults and unemployed youth. Thus, the experimental evaluation of this program was assigned to both of these categories. Most experiments were assigned to only one target group category, however. Nonetheless, the fact that some experiments were assigned to two or more target groups causes the totals in table 1 to exceed the 150 experiments in our sample.

In using the program typology, experiments that evaluated existing programs or modifications in these programs were classified on the basis on the program or program

modifications being evaluated. Thus, for example, a particular welfare-to-work experiment that tested a package of components that included training, remedial education, work experience, job search, and financial incentives was assigned to the welfare-to-work category, rather than to separate categories for each of the five program components being tested. The objective was to determine the frequency with which existing programs have been the subjects of social experiments. However, some experiments, such as the Supported Work Experiment, tested innovative programs that have never been adopted as an ongoing policy in the U.S., while others, such as the job club experiments, evaluated policies that have rarely or never existed as separate entities, except on a pilot basis (for example, job clubs are typically used as one component in a package of program components). In these cases, which are fairly infrequent and mainly pertain to older experiments, special categories were created. Some experiments were assigned to more than one program category, but not very many. However, to the extent they were, they cause the totals in table 1 to exceed the 150 experiments in our sample.

In the remainder of this section 2.0, we separately discuss each of the categories listed in table 1.

2.1 Target Group Typology

Unemployed Adults

A number of experiments have tested programs aimed at unemployed adults. As can be seen from table 1, most of these experiments have involved either training programs or the unemployment insurance program. This is to be expected, as these programs have been the subject of social experiments fairly often and unemployed adults comprise a large share of the clients of these programs. Unemployed adults have also been the targets of programs that have assessed job search and job placement programs. These experiments will be discussed in more detail later, when we examine the use of random assignment in these program areas.

Unemployed, Out-Of-School Youth

Most of the programs for unemployed, out-of-school youth that have been evaluated by random assignment have provided some combination of remedial education, vocational training, work experience gained through subsidized employment, and job search. These experimental evaluations have indicated that most of the programs that provided these services failed to increase the earnings or employment of this group. However, there are major exceptions including job clubs (see below), the Job Corps, and San Jose's Center for Employment and Training (which was experimentally evaluated as part of both the Minority Single-Parent and the JOBSTART demonstrations), all of which have been experimentally evaluated with favorable findings. The approach used in San Jose is currently being assessed by an experiment being conducted in other locations to see if San Jose's success is replicable.

There now seems to be little evidence that remedial education, vocational training, work experience gained through subsidized employment, or job search services (other than job clubs) are effective for unemployed youth who have left school. However, there may be more encouraging results when the Center for Employment and Training replication experiment is completed. Nonetheless, disadvantaged out-of-school youth are an important target population,

one that continues to suffer high unemployment and poor employment prospects. The Department of Labor will undoubtedly want to find employment-oriented programs that are effective for this population. In view of this fact, DOL may want to pursue a strategy of planned experimentation to find policy approaches that are more effective in improving young people's employment rates, job qualifications, and wages than policies tried in the past. We discuss this more specifically at the end of section 4.0.

In-School Youth

One of the experiments in this category assessed job placement services for graduating high school students and one was used to evaluate the Upward Bound Program, which encourages low income high school students to go on to college. The remaining experiments tested a variety of approaches (for example, counseling, curriculum changes, combining work with education, and financial incentives) intended to keep students who were at risk of dropping out of high school from doing so. Several of the experimental treatments did, in fact, reduce dropout rates. For example, the evaluation of the Quantum Opportunities Program—which was targeted at high school students from families receiving public assistance and provided many of the services listed above, as well as financial incentives—found that treatment group members were more likely to graduate from high school (63 percent versus 42 percent) and to enter postsecondary school (42 percent versus 16 percent).

In addition to the experiments assigned to the in-school youth category, there have been at least half a dozen welfare-to-work experiments that have tested whether financial incentives (i.e., increases or decreases in benefit payments) can be used to increase school attendance or enrollment. As discussed later, some of the tested financial incentives were found to be successful in increasing school attendance.

As indicated above, experiments that are targeted at in-school youth typically measure program effects on schooling—for example, enrollment, attendance or graduation. It would also be very useful to know whether increased schooling ultimately translates into increases in earnings. This would, of course, require a much longer follow-up period (indeed, if schooling increases, earning could be reduced in the short term), but could be done using Social Security Administration records.

Welfare Recipients

One major, if perhaps unsurprising, finding from table 1 is that around half of the experiments that met our criteria evaluated programs that were targeted at welfare recipients. Most of these programs were funded by the Department of Health and Human Services (DHHS) or state welfare agencies or both and were targeted at AFDC (Aid to Families with Dependent Children; now called Temporary Assistance for Needy Families or TANF) beneficiaries. However, a few have been sponsored by the Food and Nutrition Service, which oversees the Food Stamp Program, and several have been at least partially funded by foundations such as the Ford and Rockefeller Foundations. These programs, which are discussed at greater length below, are usually referred to as “Welfare-to-Work Programs.” Their objective is usually to move welfare recipients into employment and to reduce recipients’ dependency on cash public assistance, although they may have other goals as well. As suggested by table 1, other programs

that specifically targeted welfare recipients have also been evaluated by social experiments, including a few training programs that have been funded by DOL.

Nonworkers With Severe Employment Barriers Not Elsewhere Classified

This category covers experiments that evaluated programs that are targeted at a variety of highly disadvantaged nonworkers including the disabled, the mentally impaired, the homeless, ex-convicts, and recovering substance abusers. Although a number of work-orientated experiments have been targeted at these groups, the treatment tested by quite a few of them was limited to case management, help in finding a job, and employment-orientated counseling. However, other experiments have tested a wide variety of more intensive services including job clubs, training, vocational rehabilitation, and work experience gained through subsidized employment. Most of these experiments have involved relatively small samples. One significant exception to this is the National Supported Work Demonstration, which was conducted in the late 1970s and which targeted ex-convicts and recovering substance abusers (as well as long-term welfare recipients and young school dropouts). The tested program consisted of employment in a structured work experience program involving peer group support, a graduated increase in work standards, and close supervision. A more recent important exception is the experimental test of Project Network, a program in which disabled persons on Supplemental Security Income or Social Security Disability Insurance received employment assessments, help in developing individual employment plans, and arrangements for the services (such as employment placement, rehabilitation, and training) needed to achieve the employment plans. Although some of the experimentally tested programs that were targeted on nonworkers with severe employment barriers (including Supported Work) were not found to have positive impacts for these persons, others (including Project Network) have been found to increase their employment and earnings.

Other than the Supported Work Demonstration, there have been very few experiments that have targeted ex-convicts. However, one experiment that was conducted in the early 1970s (the Living Insurance for Ex-Offenders Demonstration) found that while job placement services were not successful for this group, financial incentives were highly cost-beneficial. These results are based on a relatively small sample, however. Given the current size of the prison population, additional, more contemporary experiments that target individuals who have recently been released from prison are probably warranted.

Currently Employed People Who Earn Low Wages

This category includes experiments that were targeted on a variety of low-wage employed workers, including migrant and seasonal workers, older workers, and former unemployment insurance claimants and AFDC recipients who found low-wage jobs. The unemployed, in addition to the currently employed, participated in the treatment tested by four of the experiments we include in this category. For example, two experiments assessing programs that help individuals start their own businesses targeted both the unemployed and the currently employed.

The treatment tested in three of the remaining experiments, which might be characterized as “post-employment experiments,” was targeted at recently employed ex-AFDC recipients. The

goal of these experiments was to keep former recipients employed, and in some cases, to help them advance to better jobs. Some combination of training, education, financial incentives, career planning, case management, and counseling was used to do this. Two post-employment experiments are complete, but neither achieved positive results. However, the third of these experiments, which has only recently been fielded, is the most ambitious of the three. This experiment, the Employment Retention and Advancement Project (ERA), is testing a wide variety of services and financial incentives at 23 different sites. In addition, a large-scale post-employment experiment, which is also called ERA, has been recently initiated in Great Britain. While similar to the U.S. ERA experiment, it puts more emphasis on financial incentives and has more varied target groups, including the long-term unemployed and low-wage single parents who have not been on welfare, as well as single parents who have received welfare benefits.

The evaluation of programs that provide post-employment services and financial incentives may be an important area for further experimentation. Moreover, there is no reason why such evaluations should be limited to programs that target the welfare population. Post-employment programs that target other groups with unstable work histories or employment in dead-end jobs could also be usefully evaluated in the United States, as they are in Great Britain. However, future post-employment experiments should probably wait until results are available from both the United States and the U.K. ERA demonstrations.

Native Americans

Native Americans, especially those who live on reservations, are eligible for a number of programs that are specifically targeted on them. However, to the best of our knowledge, none of these programs has ever been examined in a social experiment. Many reservation Indians live in relatively isolated and tightly knit communities, and thus experimental cross-over and contamination may be more likely on reservations than elsewhere. The reasons for the lack of experimentation with this target population are not really clear, however. One partial explanation is the relatively small number of participants in programs specifically targeted at Native Americans. Nonetheless, if sufficient care is given to implementing and monitoring random assignment procedures, social experimentation on Indian reservations should be feasible.

2.2 Program Area Typology

DOL Training and Education Programs

As indicated by table 1, DOL-funded job training and education programs that are targeted at unemployed adults and youth have been involved in a number of social experiments. This is to be expected, given the importance of training programs as both a labor market intervention and as a focus of nonexperimental evaluation. Some of these experiments have been multi-site and have involved quite large samples. Indeed, as discussed in the following paragraph, two of these evaluations are among the most important social experiments ever conducted. Moreover, the training component of Welfare-to-Work Programs has typically been provided by DOL-financed programs. In this sense then, DOL training programs have been evaluated by random assignment much more frequently than table 1 implies at first blush. However, the programs have been indirectly evaluated as part of a larger evaluation effort, and

the welfare-to-work evaluations focus on a comparatively narrow group of participants in DOL-funded programs.

Social experiments have frequently been used to evaluate incremental changes to existing government programs. Much more rarely, they have been used to conduct overall assessments of important and long-established government programs. However, they have recently been used in precisely this manner to evaluate two major DOL training programs: the Job Corps and the JTPA, which was the immediate precursor of the Workforce Investment Act. The findings from these experiments have been influential. For example, the evidence from the JTPA experiment indicated that the program increased the earnings of adult men and women by around a \$1,000 over the 30-month followup period, and hence, resulted in net benefit for society, but the program was found to be ineffective for youth. In response to this finding, the Clinton administration recommended, and Congress agreed, to a sizeable reduction in JTPA funds allocated to youth, while funding for adults has remained intact.

Because of the importance of job training programs and because their designs continually evolve, consideration should be given to conducting further experiments along the lines of the Job Corps and JTPA experiments. We discuss this topic further in section 4.0.

Welfare-To-Work Programs

As noted earlier, a very large number of welfare-to-work initiatives have been evaluated through random assignment experiments. Most of the evaluated Welfare-to-Work Programs, but far from all, have utilized the threat of sanctions (i.e., reductions or termination of cash benefits) to enforce participation in job search training, education, and work experience programs. Several experiments also tested changes in financial incentives to encourage welfare recipients to find and keep unsubsidized jobs. Although not indicated in the table, the target group of most of these Welfare-to-Work Programs consisted of female heads of single parent families who were receiving AFDC (now called TANF). A number of experiments also enrolled married- and unmarried-couple families if they were receiving or applying for AFDC or TANF benefits. Some Welfare-to-Work Programs evaluated through random assignment provided services that were targeted on the children in these families. A few of these provided training to teenagers, but most focused on using financial incentives to improve youngsters' school attendance. Finally, there have been a few random assignment evaluations of Welfare-to-Work Programs operated by the Food Stamp Program.

Why have there been so many experimental evaluations of welfare-to-work initiatives? One reason is that welfare programs have been highly controversial and politically unpopular. Thus, there is pressure to test new initiatives that result in additional government expenditures carefully and demonstrate that they are effective. However, the full story is more complex.¹⁶ The Omnibus Budget Reconciliation Act (OBRA) of 1981 for the first time allowed states to implement Welfare-to-Work Programs of their own design. States usually needed to obtain Federal waivers before new programs could be implemented, however. To receive Federal waivers, states had to agree to an evaluation of the Welfare-to-Work Programs that they initiated in response to the OBRA legislation. To be sure, the evaluation did not have to be performed using a random assignment design.

This began to change, however, when Manpower Demonstration Research Corporation (MDRC), a New York based evaluation firm with a strong institutional commitment to random assignment, decided to use experimental methods to evaluate the programs in several states. It raised funds through a Ford Foundation grant and other sources that permitted low-cost experimentation in a handful of states. MDRC was able to persuade some (but not all) the states it approached to allow random assignment evaluation, in part, because of the prestige of participating in a Ford Foundation project and, in part, because welfare administrators in some states wanted information about the effectiveness of the programs they implemented. Around 1987, after findings from the earliest welfare-to-work experiments became available, officials at DHHS became convinced that experimental designs were superior to nonexperimental methods. These officials were responsible for issuing waivers to the states that wanted to implement Welfare-to-Work Programs, and over time DHHS officials began to require or strongly encourage random assignment evaluations as a condition for granting waivers for state welfare-to-work plans. Thus, the large number of welfare-to-work experiments is due to Federal legislation that encouraged states to test various ways of moving welfare recipients into the workforce and the commitment of a research firm and (later) Federal officials to using random assignment to evaluate these initiatives.

A recent meta-analysis that was based on 24 random assignment evaluations of over 60 Welfare-to-Work Programs found that most of the impacts on earnings were positive but modest (\$89 per quarter when averaged over all the calendar quarters of available data); but that these impacts varied considerably among the programs. Similarly, most of the programs also reduced the receipt of AFDC, but only by a small amount (1.7 percentage points when averaged over all the programs and available calendar quarters). Again there was considerable variation across the programs.¹⁷ Cost-benefit analyses of these programs have indicated that most save the government money (i.e., reductions in transfer benefits exceed the costs of operating the program), while many (but far from all) also make welfare recipients better off (i.e., increases in earnings exceed reductions in transfer benefits).¹⁸

By providing evidence on programs that are more effective and less effective, the welfare-to-work experiments have influenced welfare policy within particular states (for example, California) and have also influenced Federal policy, particularly the Family Support Act of 1988.¹⁹ However, their influence on the Personal Responsibility and Work Opportunity Act of 1996 (the Federal legislation that resulted in arguably the most important changes in welfare policy since the 1930s) appears to have been minimal (Greenberg et al. 2003).

Other Non-DOL Training and Education Programs

This fairly large category, which includes a substantial number of experiments, is something of a catch-all. As indicated by table 1, it covers most of the target groups in our taxonomy. Moreover, the experiments assigned to this category were funded by a number of government agencies and, in some instances, by foundations. In addition, the services provided by the experimental programs varied greatly. Only around one-third directly provided training or education, although this includes the programs evaluated by most of the larger experiments. Some of the remaining experiments are not readily classified into the other categories in the program area typology. The four experiments that targeted in-school youth, for example, tested

drop-out prevention programs that were intended to increase education, rather than directly providing it. A number of the other experiments in the category provided employment-orientated counseling and case management, channeling some, but not all, of those in the experimental group into training and education.

Subsidized Employment and Work Experience Programs

Subsidized employment and work experience programs typically assign participants to jobs provided by a government, nonprofit, or (occasionally) profit-making organization. The jobs at these organizations are subsidized in the sense that the output produced by those who hold them may be less than the subsidized workers are paid. Welfare-to-Work Programs also often place some participants in work experience job, but unlike all but a few of the programs tested by the experiments assigned to this category in table 1, work experience jobs that are provided by Welfare-to-Work Programs usually do not pay a stipend.

Subsidized employment and work experience jobs are often combined with other services. In some programs, those who work at these jobs receive specific vocational skills as a result of the experience. Other programs provide less meaningful employment. However, it is anticipated that participants with little previous work experience will learn appropriate work place behaviors and this will help them when they try to move into regular jobs. There is, in fact, some experimental evidence that these programs can be effective, although some tested programs have been found to be ineffective. A majority of the experiments listed in table 1 in this category found that the programs they tested resulted in increases in earnings and employment after participants left the program.

Nonetheless, it is not apparent that more experiments should be conducted in this area anytime soon. Subsidized employment and work experience programs are not currently in vogue (although they could be at some point in the future). This is reflected by the fact that all but one of the experiments included in this category in table 1 were completed by 1990, and most were conducted considerably earlier. Additional experimentation should probably wait until the possibility of running such programs on a moderate to large scale is again seriously considered.

Trade Adjustment Assistance Program

As indicated by table 1, the trade adjustment assistance program, or possible modifications to that program, has never been evaluated through random assignment. The program itself cannot be tested experimentally because eligibility for it is an entitlement. In principle, however, there is little reason why random assignment could not be used to evaluate modifications to the program. Several of the services provided through this program, such as job search assistance, retraining, and cash benefits, are common to other programs that have been subjected to random assignment. One reason random assignment has not been used for this purpose is the program's relatively small size. To obtain a sample that is of adequate size for an experiment, it would have to be conducted in a state with a sufficiently large program. There are only a few states that meet this requirement, and they would have to be willing to have an experiment conducted in their state. In the one previous attempt to test an enhanced trade adjustment assistance service, none of the states with sufficiently large programs were interested.

Entrepreneurship and Self-Employment Programs

Two previous experiments have tested the effects of providing aid to individuals who want to start up small businesses. One was very small and had serious design problems. The other, the Self-Employment Enterprise Development Project, which was targeted at new UI claimants, produced generally positive effects and the tested treatment was found to be cost-beneficial at both of the experimental sites. A third experiment, Project GATE (Growing America Through Entrepreneurship), has very recently been initiated at seven sites in three states.²⁰ Members of the Project GATE treatment group are provided training and technical assistance in running a business, including help in applying for business loans.

Presumably, there have been only a small number of experiments that test the effects of the government providing business start-up services because relatively few individuals want to be self-employed. Nonetheless, Congress has authorized states to operate Self-Employment Assistance Programs for UI claimants, and other self-employment programs are operated by community action groups and community development corporations. Thus, it seems worthwhile to use random assignment evaluations to assess whether these programs are effective and to determine the array of services that are most useful. However, further experimentation in this area should probably wait until findings from the ambitious Project GATE evaluation become available.

Unemployment Insurance (UI)

Table 1 indicates that social experiments have been used a number of times to assess the merits of proposed changes to the UI Program. Several of these experiments involved evaluations of various combinations of changes in the claims process, stricter enforcement of the UI work test, the utilization of case management, and enhanced placement services for claimants. In addition, there are four well-known experiments that tested the effects of a variety of reemployment bonus schemes. As might be expected, the major target group of the UI experiments was the major UI claimant group—unemployed adults.

The UI experiments were largely motivated by efforts to discover methods of reducing the cost of the UI Program through either administrative simplifications or getting claimants back to work sooner. Although findings from these experiments have entered the policy debate over the UI Program, the tested policy changes have not been adopted (Greenberg et al. 2003).²¹ This does not mean the experimental findings were not used, however. Lack of implementation of the tested policies is due, in part, to the fact that results from the experiments have not been especially supportive of them. For example, although findings from the four reemployment bonus experiments indicated that such schemes reduce the length of unemployment spells, they also suggested that the costs of the bonuses more or less offset the resulting savings in UI claimant payments.²² Furthermore, the experiments were used in other ways. For example, data from the UI reemployment bonus experiments have aided the development of profiling schemes used nationally in administering the UI Program, as well as other programs. In addition, the successful implementation of random assignment in running the reemployment bonus experiments appears to have motivated DOL to run other types of experiments (Greenberg et al. 2003, 187-189).

As new changes in administering the UI Program are proposed, consideration should be given to subjecting them to random assignment evaluation. For example, profiling, which is now used nationally to determine which UI claimants should receive mandatory reemployment services (e.g., structured job search, employment counseling, and retraining), has to the best of our knowledge been tested through random assignment only once, and then only in one state (Kentucky).

In addition, the Bush administration has introduced legislation that would provide certain UI claimants (those that profiling indicates are most likely to exhaust their benefit entitlement) with reemployment account funds that they can use to purchase reemployment services, receive as income maintenance if they exhaust their UI benefits, and receive as reemployment bonuses if they obtain employment within 13 weeks. Because the findings from the reemployment bonus experiments were somewhat mixed and the proposed program is expensive (an estimated \$3.6 billion over two years), consideration should be given to testing it on a pilot basis through random assignment before adopting it nationally.

Even if the policy is adopted as proposed by the administration, there is still a good case for testing planned variations of proposal within one or more social experiments. As discussed in section 4.0, random assignment could be used to determine the effects of varying the duration of program eligibility and the weekly benefit amount. In addition, it could be used to determine the effectiveness of alternative service packages to help UI recipients who will be offered reemployment account funds.

Job Search and Placement Programs

Job search and job placement have frequently been embodied in larger programs that also provide other services. For example, they are almost always a component of Welfare-to-Work Programs and training programs. However, as indicated in table 1, there have been a number of experiments that have evaluated job search or job placement in isolation. These experiments were targeted at a variety of groups. Some of them focused on programs that provide help with resume writing and employment counseling. Others have also provided placement services. A number of the experiments that are assigned to this category in table 1 tested the effects of job clubs on different target groups, mostly during the 1970s and early 1980s.

In job clubs, participants learn various job search skills and techniques and engage in supervised job search. These experiments all found that the effects of job clubs on employment and other outcomes were positive, and sometimes surprisingly large. Because job clubs are relatively inexpensive to operate, the experiments also implied that they are cost effective. Although these experiments were small, they have been quite influential. Job clubs are now widely used as one component in training and welfare-to-work in the U.S. and elsewhere. The early job club experiments appear to have produced sufficiently definitive findings that further random assignment experimentation is probably not necessary.

Labor Exchange Programs

Although there have been numerous social experiments that have evaluated enhanced job search and placement services, either separately or as one of several program components, social

experimentation has been rarely used to assess the United States Employment Service or possible modifications in labor exchange services. One exception is a 1976 experiment that found that whether or not employment counseling was provided clients had little impact on their duration of unemployment, wages, and a number of other outcomes.

Random assignment of individual Employment Service clients is probably no longer feasible, given the transformation of the Employment Service into the One-Stop Career Center delivery system. Indeed, the very essence of this transformation is to provide easy access to all the services provided at One-Stop Career Centers, probably making establishment of a control group infeasible. In principle, however, it would be possible to randomly assign the One-Stop Career Centers themselves to different treatment groups and vary these groups in terms of staff size and quality, software and equipment, or services provided. As discussed in section 4.0, however, it is not clear that such an experiment would be very practical.

Minimum and Living Wage Programs

As shown in table 1, minimum wage and living wage programs have not previously been subject to social experimentation, presumably because it would be very difficult to implement random assignment in the context of these programs. For example, it would be interesting to compare the employment status and living standards of workers who were randomly assigned different minimum wages, but it is not apparent how in practice the minimum wage could be allowed to vary among otherwise similar individuals. On the one hand, special legislation would be required to pay some workers less than the statutory minimum, and it seems unlikely that Congress would be willing to pass such legislation. On the other hand, it seems unlikely that employers would be willing to pay some workers at the statutory minimum wage and other similar workers at a wage that is above this level, and requiring them to do so would probably not be politically acceptable.

Demand Side Subsidies

Several experiments conducted in the early 1980s tested using subsidies and tax credits that were intended to induce employers to hire disadvantaged workers. Interestingly, in two of these experiments, it was the employers rather than the disadvantaged workers who were randomly assigned. All of these experiments, both the two that randomly assigned individuals and the two that randomly assigned workers, have indicated that demand side subsidies are ineffective. Thus, these experiments provided important information. Because of the consistently negative results, however, additional experiments in this category may not be warranted unless Congress or the administration wish to consider new subsidy schemes that differ considerably from those tested in the past.

One reason demand-side subsidies have been found ineffective in the past is that many employers are reluctant to hire job applicants who have identifiable characteristics suggesting they will be unproductive or undesirable employees. Unfortunately, when an employment agency identifies a job applicant as eligible for a targeted earnings subsidy, it also identifies the applicant as someone who may be an unproductive or undesirable employee. It may be possible to design more effective demand-side subsidies if such subsidies are only provided to job applicants whose disabilities would become known to employers even in the absence of the

government subsidy. For example, blind, physically disabled, and aged job applicants have disabilities that would become known to employers in the normal hiring process. Ex-offenders recently released from prison also have a disability that would soon become known to employers. It may make sense to test demand-side subsidies that are targeted on groups with these kinds of obvious disabilities. Demand-side subsidies tested in past experiments usually target subsidies on populations such as welfare recipients where the job applicant's disadvantage would not become known to an employer unless an employment agency reveals this disadvantage when advertising the applicant's eligibility for an earnings subsidy.

Supply Side Subsidies

This category pertains mainly to experiments that test financial incentives that are intended to induce individuals to either find employment or continue working once they have jobs or both. As mentioned above, such incentives have previously been targeted at welfare recipients and UI claimants in programs that have been evaluated through random assignment. However, these experiments are not counted in table 1 as supply side subsidies, but are instead listed under other headings (e.g., welfare-to-work and unemployment insurance). Thus, financial incentives have been tested considerably more often than the table indicates, although most often as one component of a program that also provides various services.

On balance, findings from various social experiments seem to indicate that financial incentives can increase employment and earnings, albeit at some cost to the government. The reemployment bonus experiments for unemployment insurance claimants, which were discussed earlier, provide an important example of this. However, when financial incentives are combined with services, as they typically are, it is usually difficult to isolate the separate effects of job placement services and financial incentives. Fortunately, there have been a few experiments that have randomly assigned some individuals to a program that provided only financial incentives and others to a program that combined financial incentives with services. Findings from these experiments imply that financial incentives are more effective when combined with services. Thus, experimental evaluations of financial incentives that are provided in the absence of services may understate their potential effectiveness.

Financial incentives can vary along a number of different dimensions—for example, their size, the conditions that must be met to be eligible for them, the length of eligibility, and the services that are provided with them. Further random assignment experiments that test financial incentives should attempt to find the most optimal combination of these parameters. This topic is further discussed in section 4.0.

Loans and Subsidies for Education and Training

There have been four experiments that have tested whether financial aid schemes can encourage additional education. The treatment tested in one of these experiments, which was targeted at individuals who left the welfare system for employment, did not appear to have positive effects. However, the experimental treatment was not implemented as designed and the experiment was terminated a year earlier than planned. The second experiment had severe implementation and design problems, but did suggest that scholarship offers have positive effects on the college enrollment of low income youth. The third experiment, which was targeted at low

income adults in both single- and two-parent families, found that a 100-percent tuition subsidy for training or education of the individual's choice resulted in increases in schooling for all the study groups and a 50-percent subsidy increased schooling among single parents. Surprisingly, however, the subsidies appeared to reduce earnings even after the training was completed. Results from the fourth experiment, which is testing the effects of individual training accounts for unemployed and displaced workers in six different states, will not be available for several more years. Individuals in the experimental group are free to spend the funds in their training account on any state approved training program and on related expenses approved by local counselors.

There have also been at least six so-called “learnfare” experiments that have examined whether positive or negative financial incentives can increase the amount of pre-college education received by children and teenagers in AFDC families. In addition, the Teenage Parent Demonstration reduced the welfare grant of teenage AFDC mothers who did not participate in training or education for at least 30 hours a week. Depending on what the tested financial incentives rewarded or penalized, outcomes were measured in terms of school enrollment or attendance or receipt of a high school diploma or equivalent. Four of the tested treatments (including the one tested by the Teenage Parent Demonstration) increased the amount of education received, sometimes appreciably, and three did not. Of the four more successful programs, two resulted in small to modest positive effects on employment and earnings, one had no impact on these outcomes, and one did not examine them.²³

This would appear to be an area where more social experiments could usefully be conducted. Particular consideration should be given to using random assignment to test the usefulness of financial aid to encourage college education among low income youth, given the absence of definitive experimental findings for this group. Additional experiments that test policies that are targeted at adult unemployed workers should probably wait until after findings from the individual training account experiment mentioned above become available.

Federal Bonding Program

The Federal Bonding Program provides bonds to employers to protect them against theft, forgery, larceny, and embezzlement that might result from hiring so-called “at-risk job applicants”—for example, ex-offenders, those dishonorably discharged from the military, recovering drug and alcohol abusers, and low income persons without a work record. The bonds are issued by the Travelers Property Casualty Insurance Company and provided to employers free of charge for the first six months they employ covered individuals. Only 40,000 workers have been covered by the program since it was established in 1966.

The program has never been the subject of a random assignment evaluation, possibly because it is small and not well known, but it possibly should be. If bond insurance coverage appears to increase employment, consideration could be given to making more individuals with serious employment barriers eligible or at least more highly publicizing the program. A social experiment could be readily conducted by randomly assigning some eligible individuals to a control group that is made temporarily ineligible for the bond or by varying the bond amount or the length of time the bond remains effective by random assignment.

Table 1. Experiments by Target Groups and Program Areas

	Unemployed adults (excluding experiments targeted only at welfare recipients)	Unemployed, out-of-school youth (excluding experiments targeted only at welfare recipients)	Welfare recipients	Non-workers with severe employment barriers not elsewhere classified (includes the disabled, mentally impaired, the homeless, ex-convicts, recovering substance abusers)	In-school youth	Currently employed people who earn low wages (includes migrant and seasonal workers, older workers, and unemployed workers and former UI claimants and AFDC recipients who have found jobs)	Native Americans	<u>Program totals</u>
DOL Training and Education Programs	6	7	3	0	4	0	0	<u>19</u>
Welfare-to-Work Programs	0	1	67	1	0	0	0	<u>69</u>
Other Non-DOL Training and Education Programs	4	2	0	8	4	5	0	<u>23</u>
Subsidized Employment and Work Experience Programs	0	4	4	3	0	0	0	<u>11</u>
Trade Adjustment Assistance	0	0	0	0	0	0	0	<u>0</u>
Entrepreneurship and Self-Employment Programs	4	0	1	1	0	2	0	<u>8</u>
Unemployment Insurance	10	0	0	0	0	0	0	<u>10</u>
Labor Exchange Programs	1	0	0	0	0	0	0	<u>1</u>
Job Search and Placement Programs	8	4	2	6	1	0	0	<u>21</u>
Minimum and Living Wage Programs	0	0	0	0	0	0	0	<u>0</u>
Demand Side Subsidies	0	2	2	0	0	0	0	<u>4</u>
Supply Side Subsidies	2	1	0	1	0	2	0	<u>6</u>
Loans and Subsidies for Education and Training	2	1	0	0	0	2	0	<u>5</u>
Federal Bonding Program	0	0	0	0	0	0	0	<u>0</u>
<u>Target Group Totals</u>	<u>37</u>	<u>22</u>	<u>79</u>	<u>20</u>	<u>8</u>	<u>11</u>	<u>0</u>	<u>177</u>

3.0 Lessons from Past and Current Experiments

In this section, we examine some lessons from past experiments. We are particularly interested in uncovering lessons that would shed light on the kinds of questions where experimentation has been most useful. Have past experiments provided useful knowledge for improving policy? Under what conditions has that knowledge actually been applied? Did the results of experiments offer a reliable guide to impacts when tested policies were adopted? We are also interested in drawing practical lessons about how the design of workforce evaluation experiments can be improved.

In section 1.0 it was argued that experiments are preferable to other kinds of policy evaluation when the existing, nonexperimental evidence about policy effectiveness is subject to doubt and when a random assignment trial can reduce our uncertainty about the size and direction of program impact in a cost-effective way. Section 2.0 and the appendices briefly describe and analyze the large number of workforce program experiments that have been conducted to date. These experiments can help us answer the questions posed in this section.

3.1 Feasibility, Cost-Effectiveness, and Time Costs

The long list of experiments offered in appendix A of this report demonstrates that randomized trials are feasible for workforce policy evaluation. In the view of these authors, the overwhelming majority of completed experiments were reasonably successful in maintaining the integrity of treatment and control groups. Almost all experiments obtained internally valid estimates of the treatment-control difference in the enrolled sample. The external validity of their estimates is harder to assess. Except for the experiments that examined the impact of ongoing programs, most experiments have tested policies that have never been implemented in precisely the form that was tested in the experiment. In the discussion below, it will be considered whether the welfare-to-work experiments provided useful information for predicting the effects of the welfare reform policies adopted by the Federal and state governments over the second half of the 1990s.

The history of workforce evaluation experiments suggests randomization can be costly. Details about their financial cost can be found in other sources.²⁴ For a number of experiments, one of the most striking costs involved is the time consumed in implementing and completing the experiment and publishing a final report. The Seattle-Denver Income Maintenance Experiment first enrolled participants in October 1970, completed benefit payments to the last enrollees in August 1977, and published its final report in May 1983, more than a decade after enrollment began. Fortunately, interim findings from the study were available by 1977 when the Carter administration began considering alternative reform plans for cash public assistance and public employment, and the interim findings were used both by supporters and opponents of the administration plan to analyze the costs and benefits of alternative reform plans.

Most experiments have been completed in much less time than the Seattle-Denver Experiment. The Washington State Reemployment Bonus Experiment began enrolling UI recipients in February 1988 and issued a final report in 1992. Some experiments take even less time to yield results. The Dayton Wage Voucher Experiment was designed in the middle of

1980, began to enroll eligible job seekers in December 1980, and completed enrollment in May 1981. Preliminary findings from the experiment were known to DOL by June 1982.²⁵

The wide differences in the time it takes to design and complete an experiment are due to a couple of factors. One is the complexity of the experimental design. The Seattle-Denver Experiment tested 11 financial incentive plans in combination with 4 workforce counseling and educational voucher plans. In addition, several of the financial plans were tested for both 3 years and 5 years. The experimental designers created an elaborate design to minimize the cost of benefit payments while maximizing the statistical power for estimating crucial behavioral parameters. In contrast, the Washington Reemployment Bonus Experiment tested a smaller number of financial incentives, and the Dayton Wage Voucher Experiment tested only two variants of a wage voucher. The experimental designs were simpler and much easier to implement. Many experiments test only a single treatment variant. Another difference arises because of differences in the time interval required to observe enrollee outcomes. In the Seattle-Denver Experiment, analysts were interested in measuring enrollees' labor supply response to a variety of financial incentives. To be meaningful, measurements of this kind of response require financial incentives be maintained and labor supply responses to be monitored over several years. In the Reemployment Bonus and Wage Voucher Experiments, analysts were principally interested in the impact of financial incentives on the duration of unemployed workers' job search and receipt of public transfer benefits. These kinds of effects can be reliably measured in the first year after an unemployed worker is enrolled.

This contrast between the two kinds of experiments highlights a simple point. The duration of an experiment depends critically on the nature of the behavioral response that policymakers wish to observe. If they are interested in assessing the impact of a treatment on the duration of job search or the attainment of a particular educational credential, such as a GED diploma, the entire response to the tested treatment can be measured in a short experiment. More frequently, however, analysts and policymakers are interested in responses that may take several years to observe. In many workforce training programs, we are interested not only in whether an eligible person takes advantage of training and obtains a credential, we also wish to know whether the extra training helps the person become or remain employed and earn good wages. In light of this interest in employment and earnings, many experiments do not last long enough.

Consider a tested treatment that succeeds in persuading enrollees to make greater investments in education or skill acquisition. Many treatments that encourage disadvantaged job seekers to invest in schooling or training divert them, at least temporarily, from paid employment. In comparison with job seekers in the control group, those in the treatment group take longer to find jobs. They earn less income in the first 6 to 18 months after they are enrolled than job seekers in the control group. This pattern is clearly visible in the earnings response to the Job Corps Program measured in the recent Job Corps Experiment.²⁶ If data collection in an experiment ends prematurely, analysts may not be able to observe whether the initial drop in treatment-group earnings is offset by later gains that result from better training or education. The Job Corps Experiment shows substantial earnings gains in the seventh through the sixteenth calendar quarters after participants were enrolled in Job Corps, but these gains alone are not large enough to offset earnings reductions during the first 6 quarters after enrollment plus the direct administrative costs of providing Job Corps services to enrollees. If analysis is restricted to the

first 4 years after young people are enrolled in Job Corps, the program has social costs which outweigh the measurable social benefits attributable to the program. The final report of the Job Corps study shows that people enrolled in the treatment group continued to earn higher wages than people enrolled in the control group during the last calendar quarter for which interview information is available. This means that the initial followup period in the experiment was too short to allow us to infer reliably whether the costs of the program exceed the measurable benefits.²⁷ (The Job Corps evaluators are collecting followup information from Social Security Administration earnings records which may allow them to obtain a more definitive assessment of the long-term earnings gains attributable to Job Corps.)

A number of experiments have examined methods to encourage public assistance recipients to leave the welfare rolls and become employed. One tested treatment is to encourage or require welfare recipients to enroll in education or training programs. A much more common treatment is to require them to participate in job search assistance programs and use the threat of sanctions to encourage recipients to accept private-sector jobs as quickly as possible. The first approach can initially depress enrollees' employment and earnings as they receive classroom training or some other kind of instruction. The second can initially boost enrollees' earnings if it actually succeeds in pushing welfare recipients into jobs. Most experimental evaluations limit followup data collection to the first 3 or 4 years after the treatment is introduced. This evaluation period makes it difficult to measure long-term earnings gains that may follow extra investments in education or training. If extra investments in schooling or training do not boost enrollees' earnings until the second or third year after training is completed, the short duration of the follow-up data may limit the ability of researchers to compare the two policy approaches. On the other hand, if the followup data show no detectable impact of training investment at the end of a three- or four-year followup period, it seems unlikely that the impact will ever be large. At a minimum, researchers should identify an inexpensive source of followup data to determine whether any earnings gains measured at the end of the planned followup period persist beyond that period. The experience of the Job Corps study suggests that such information is often crucial for performing an accurate benefit-cost analysis of the tested treatment.

Our discussion suggests that the sometimes lengthy delays in obtaining final results from a randomized trial are frequently unavoidable. A nonexperimental study of the same treatment based on the same employment and earnings information would take essentially as long as the experimental study. Assuming that a nonexperimental analysis of Job Corps data could provide unbiased estimates of the program's impact on earnings, the nonexperimental study would also have inconclusive earnings estimates after 16 quarters of followup data collection. Additional followup information beyond 16 quarters is needed to determine whether the Job Corps treatment passes a benefit-cost test, and this is true whether the impact is estimated with experimental or nonexperimental methods. When experiments find that the impacts of the treatment last up to the full length of the original followup period, there is a strong case for finding resources to continue follow-up data collection. Collecting additional followup data often represents the least costly method for reducing uncertainty about the net value of the tested treatment.

Even where an experiment will take much longer than a nonexperimental study, it is often the case that the policy question at issue will remain relevant for a long time. If the

increased reliability of the result from an experiment is significant, policymakers should often prefer to wait longer for the more reliable experimental result than to rely on the less reliable evidence from nonexperimental studies.

3.2 Have Experimental Results Influenced Workforce Policy?

As noted earlier, the simplicity of experiments often provides major advantages in making results understandable and convincing to other analysts and policymakers. This advantage has sometimes been decisive in influencing policymakers when nonexperimental studies have had very little weight in their decisions. When policymakers understand and are persuaded by the findings of a simple experiment, they can be induced to make changes in public policy. Section 1.0 mentioned the impact of the JTPA experiment findings on funding allocations in the JTPA program. The estimated effect of JTPA services on adults enrolled in the program was beneficial. In contrast, researchers found little evidence of a beneficial impact on youth. Some evidence indicated that youths enrolled in the program had worse outcomes than youths enrolled in the control group. The finding led directly and quickly to a reallocation of JTPA program funds away from the group where no benefit was found.

In contrast, the series of reemployment bonus experiments conducted in the States of New Jersey, Illinois, Pennsylvania, and Washington did not have the impact on public policy that was originally anticipated. All experiments found that the offer of a bonus reduced the duration of unemployment spells, although in a couple of experiments the reduction was small. Several of the bonus plans tested in the experiments were found to reduce unemployment spells and increase short-term earnings by enough to justify the cost of the bonuses and the management expense of administering them. Nonetheless, reemployment bonuses have not been adopted as a component of the UI program.²⁸ Both Federal and state policymakers showed little interest in adopting this innovation, although four careful randomized trials indicated the policy could modestly reduce the average duration of unemployment spells, at least in a small-scale study. One reason is that the experimental treatment failed to produce cost savings for the UI system. In addition, decisionmakers may have been uncertain whether the findings of a small-scale trial could be applied to a similar policy implemented on a national or statewide basis. As discussed earlier, this kind of incentive can have different effects if the bonus is offered to every new UI claimant compared with the situation in which it is offered to only a few claimants in a local labor market. In a careful reanalysis of the reemployment bonus experiments, O'Leary, Decker, and Wandner (1998) examined the likely effects of bonuses if they were targeted on workers who face the highest probability of exhausting their UI benefits. They found that targeting can improve the cost-effectiveness of bonuses, but even stringent targeting would not guarantee that the bonus offers would be cost effective from the point of view of nonrecipient taxpayers.²⁹ For whatever reason, the findings from the bonus experiments have been used mainly to help craft policy proposals that have so far failed to be adopted by Congress or state legislatures.

3.3 Welfare-To-Work Experiments

In the 1980s and 1990s, a number of states introduced innovative plans to encourage public aid recipients to obtain job training, search actively for jobs, and accept paid employment. A principal goal of these plans was to increase the earned income of current and former welfare

recipients and reduce their dependency on cash public assistance. Several states subjected their policy innovations to rigorous evaluation using randomized trials. These evaluations often revealed the tested programs had significant impacts in boosting enrollees' earnings and reducing their cash assistance benefits. Many observers believe the cumulative effect of the experimental findings was to strengthen the case for dramatic welfare reform. Certainly the experiments offered evidence that the most zealous work-oriented reforms often yielded the biggest increases in enrollees' earnings and biggest reductions in AFDC participation.

Most informed observers would also agree that changes in state and Federal welfare policies between 1994 and 1998 have dramatically altered the cash public assistance system for U.S. families containing children. In 1996, the Federal Government abolished the main social assistance program for indigent adults with children and replaced it with Temporary TANF. The new Federal program placed pressure on all states to adopt aggressive policies to curtail assistance benefits to poor parents who are capable of working. The head of each family on social assistance is required to work within two years after assistance payments begin. Work-hour requirements are stringent, and states face penalties if they fail to meet them. The new Federal law stipulates that the great majority of families may receive benefits for no longer than five years, it and permits states to impose even shorter lifetime limits on benefits.

The new policies adopted by states mimic many of the policies tested in the 1980s and 1990s welfare-to-work experiments. First, they require most nondisabled adult aid recipients to actively seek work or face reductions in their monthly aid payments. Second, they provide job search services to help recipients find work. Third, most states have imposed time limits on recipients' eligibility for cash assistance benefits. (However, this important element of the 1996 Personal Responsibility and Work Opportunity Reconciliation Act was not tested in any of the experiments conducted before the act was passed.) More rarely, recipients are offered training to increase their employability. Many states have also liberalized their earnings disregards, allowing recipients to keep more of their monthly benefits when they begin to work and earn wages. Most of these policy innovations were tested in one or more state-sponsored randomized trials. Although no rational person could claim that the findings of the welfare-to-work experiments caused states and the Federal Government to overhaul U.S. welfare policy, it is fair to say that the positive findings from so many experiments provided encouragement to advocates of reform and weakened opposition to reform among people who were strong supporters of the traditional system. Almost all participants in the debate over welfare reform acknowledge that the findings from the welfare-to-work experiments provided a major impetus for reform.

Did the experiments provide a useful or accurate prediction of the response to reform? The experiments were useful in predicting the direction of response. Employment rates and earnings have risen significantly among women who had a high probability of receiving benefits under the old welfare system. Some of this increase is undoubtedly traceable to reform. The experimental findings also influenced state policymakers in their choice of reform plans. The experiments generally showed that work-oriented reforms that aimed to push aid recipients directly into employment produced bigger earnings gains and larger caseload reductions than reforms that provided recipients with access to education or job training. Most states adopted plans that encourage immediate employment and discourage recipients from making major investments in education or general training.

It is less clear whether the experiments provided an accurate prediction of the overall size of response to reform. Along with the strong U.S. economy, the new welfare law helped produce an unprecedented drop in the nation's social assistance rolls. After reaching a peak in 1994, the number of families collecting cash public assistance for children dropped almost 3 million—or over 50 percent—during the next 6 years. No drop of this magnitude had occurred in the previous 40 years. The sharp decline in the assistance rolls from their peak in 1994 is at least partly due to state-level reforms that began even before Congress passed the Federal reform law in August 1996. The decline is also due to changes in the Earned Income Tax Credit (EITC) that greatly increased the amount of earnings supplementation available to low-wage workers. The increased generosity of the EITC after 1993 combined with much tougher Federal and state work requirements has contributed not only to a decline in the U.S. social assistance rolls, but also to a large jump in labor force participation and employment among lone parents, especially single mothers. The rise in the employment rate of not-currently-married mothers was probably also due to changes in social norms, which affected attitudes toward work among single mothers just as the same trends earlier affected attitudes toward paid employment among married mothers.

There has been a dramatic change in labor force behavior of the group most likely to receive assistance payments—never-married mothers who live with their own children under 18. The employment rate of never-married mothers has risen sharply relative to that of married mothers who live with their spouse. The jump began in 1994. Before that year the employment rate of never-married mothers remained relatively constant from the late 1970s through 1993. In contrast, the employment rate of married mothers living with husbands rose steadily over the period up to 1994. Starting in 1994, the employment rate of unmarried mothers began to rise sharply. There is no evidence the employment rate of married mothers increased by a comparable amount. The unemployment rate of never-married mothers also fell 8.4 percentage points (43 percent) between 1994 and 2000. A much tighter U.S. job market, the liberalization of the EITC, new welfare-to-work reform programs at the state level, and the 1996 Federal welfare reform produced major changes in the labor market behavior of unmarried mothers.³⁰

The response estimates in the welfare-to-work experiments showed increases in employment and reductions in welfare dependency, but the changes were nowhere close to those experienced by single mothers in the second half of the 1990s. It is impossible to know whether the experiments provided an accurate or wildly inaccurate prediction of the response to reform. First, as noted earlier, none of the experiments tested precisely the combination of reform plans that were later implemented by the states. On average, states may have implemented plans containing stronger incentives for welfare recipients to find work or leave the rolls. This seems a bit unlikely, because some of the welfare-to-work experiments tested plans with very strong incentives.

A second possibility is that the dramatic increase in single-mothers' employment and huge reductions in AFDC/TANF caseloads was caused by other changes in the social welfare environment and job market. We have already mentioned the liberalization of EITC tax incentives, which should have encouraged some single parents to enter employment. In addition, the overall unemployment rate fell to its lowest level in three decades during the late 1990s. (However, single mothers' employment rates remained high and welfare caseloads continued to fall even after the country entered recession in 2001.) Interestingly, researchers who use

nonexperimental methods to analyze panel data for the 50 states cannot agree on the weight to attach to welfare policy changes, EITC liberalization, and the local unemployment rate when evaluating the decline in welfare caseloads since 1994. Some researchers conclude that state welfare reform policies played only a modest role in the drop in caseloads and increase in single mothers' employment rate. Most of the decline is due to tighter job markets, more generous earnings supplementation under the EITC, or other as yet undetermined factors. If this interpretation is true, the findings from the welfare-to-work experiments are perfectly consistent with the response to welfare reform we have observed over the past decade. We are not certain this conclusion is correct. A few researchers find that state welfare reform policies played a bigger role in reducing caseloads and boosting single mothers' employment rates. What this literature shows is that that nonexperimental evidence is not conclusive in explaining the caseload reduction and employment-rate increase. It also shows the weaknesses of nonexperimental methods in inferring the impact of policy differences and policy changes on aggregate behavior.

4.0 Some Suggestions for Social Experiments

This section considers potential future social experiments that DOL might sponsor. As such, it is the key section of the report. For purposes of the section, we have developed an "A-List" and a "B-List" of potential experiments. The experiments in the first list should, in our judgment, be of higher priority to DOL than those in the second list. Our judgment is based on both what we view as DOL's research priorities and what our review of previous and ongoing experiments in sections 2.0 and 3.0 suggests about where gaps in knowledge might be usefully filled through experimentation. We also attempted to take technical and political feasibility into account. Thus, we indicate when experiments on the "B-list" are likely to be especially difficult to implement. In our discussion of the experiments on the "A-list," we briefly flush out our ideas on how each might be designed. In doing this, we indicate the potential value of the experiments if the design we outline is used and some of the limitations of the design.

Many of the suggestions for social experiments that appear below were extracted from a very useful DOL report by Burt S. Barnow and Daniel B. Gubits.³¹ Several others were obtained from DOL reports by Steve Bell and a DOL Expert Panel.³² In some cases, the authors of these three reports specifically recommended that an experimental evaluation be carried out. In other instances, they simply indicated that evaluations should be conducted on particular topics without specifically stating whether random assignment should be used. A few other suggestions were obtained from colleagues. Finally, a need for some of the experiments listed below was suggested to us by our review of previous and ongoing experiments in section 2.0.

4.1 Higher Priority Experiments

An Ongoing WIA Experiment

The evaluators responsible for the National JTPA Study have suggested that the program be continuously monitored by randomly assigning a small proportion of program applicants (say 5 percent) in a large number of sites (they suggest 50) to a control group on an ongoing basis (Orr et al., 230-231).³³ They argued that this would reduce the very long lag that usually exists

between initiating a random assignment evaluation of employment and training programs and the availability of findings (6 years in the case of their study). They also assert that by minimizing the number of program applicants assigned to the control group, this design would reduce resistance to random assignment at the local level. We strongly agree with both arguments and suggest that such an approach be seriously considered for WIA programs. A continuous experiment of this sort would also allow policymakers to see how program impacts change as new program innovations are introduced and as the economy and the characteristics of the participant population changes. Although the sample sizes at the individual sites would probably be too small to determine impacts at that level, impacts that were aggregated over all the sites would be statistically reliable if the number of sites was sufficient. Moreover, with enough sites, multilevel modeling could be used to determine how impacts vary among sites by differences in site participant characteristics, program designs, and local labor market conditions.³⁴

A Performance Standard Experiment

There has long been controversy over performance standard systems that are used in administering employment and training programs, as well as other government programs. For example, Barnow and Gubits (pp. 57-61) discuss studies that found that while administrators respond to performance standards, the standards sometimes produce perverse incentives. They also describe evidence that indicates that performance which is measured using the standards set in the JTPA performance management system is only weakly related to actual program impacts that were measured in the National JTPA Study. Finally, they mention strong dissatisfaction at the state and local levels with the current WIA performance management system.

Performance standard systems can vary in numerous ways,³⁵ but the choices made in designing these systems have inevitably been based mainly on judgment. It seems apparent that some variants are better than others, but little or no evidence exists as to which. However, such evidence could be obtained by coupling the ongoing WIA experiment described above with a second experiment that would compare the effectiveness of two or (possibly) three performance standard variants.³⁶ The performance standard experiment would be conducted by randomly assigning the alternative performance standard systems to be tested to the sites participating in the ongoing WIA experiment. Performance measured under each of the tested system would then be compared to estimates of actual program impacts on earnings obtained from the ongoing WIA experiment. The superior system, if any, would presumably be the one that produces the strongest relation between measured performance and program impacts. Because a relatively small number of sites would be involved and randomization may therefore be incomplete, it would be useful to use multilevel modeling in estimating this relationship in order to control for other factors that may influence site-impact levels. Because of the continuous nature of the ongoing WIA experiment and because the participating sites would probably be rotated from time-to-time, the performance standard experiment could be repeated. Overtime, this would allow evidence to accumulate about quite a few different systems, and hopefully ultimately result in an improved performance management system.

Assessing Alternative Service Strategies

Both the DOL Expert Panel (p. 167) and Barnow and Gubits (pp. 42 and 48) state that it is important to determine the effectiveness of alternative training service strategies. We agree. In the National JTPA evaluation, separate impact estimates were obtained for three alternative service packages to which program applicants were assigned. As Orr et al. point out, however, many individuals in the JTPA evaluation sample crossed over to a service package that differed from the one to which they were assigned. It would obviously be better to have impact estimates for services actually received, rather than for services assigned. Moreover, assignment to the different service packages was not by random assignment. As a result, those assigned to different strategies tended to differ in characteristics, as well as in services received.

The National Evaluation Welfare-to-Work Strategies (NEWWS) suggests a different approach. In three of the NEWWS sites, welfare recipients were randomly assigned to one of three separate program groups: a group that was required to participate in a program that relied upon a human capital approach, a group that was mandated to participate in a program that emphasized a work-first approach, and a control group. This design allowed a direct comparison of the effectiveness of human capital (that is, one that emphasized investments in remedial education and training) and work-first approaches, as well as a comparison of each approach with the control regime (that is, the status quo). A similar experimental design could be used at a few WIA sites that are interested in comparing alternative service strategies and, thus, willing to participate in an experiment.

One potential candidate for a head-to-head comparison is classroom training in occupational skills versus on-the-job training (OJT). However, there is an important difference between WIA programs and the Welfare-to-Work Programs evaluated in NEWWS: participation is voluntary in the former, but mandatory in the latter. Thus, individuals who are not randomly assigned to the training package they prefer may simply drop out of training. Another difficulty is that it is typically difficult to recruit employers who will offer OJT positions to disadvantaged job seekers. Even though previous evaluations and random-assignment experiments suggest that OJT is effective in boosting participants' employment and earnings, program operators have only limited ability to assign enrollees to this kind of program because few employers are willing to offer OJT positions. Under these circumstances, comparing the effectiveness of classroom training with OJT within the context of a small-scale randomized trial could be misleading. While it is possible to offer an almost limited number of classroom training positions to program enrollees, the number of potential OJT positions is likely to be much smaller.

Another possible candidate for a head-to-head comparison, one that is less subject to the problems just mentioned, is short-term training versus more intensive, longer-term training. Barnow and Gubits (p. 38) suggest that it would be useful to test experimentally to see if additional expenditures on training results in additional returns.

An Experimental Evaluation of Retraining and Job Search Assistance for Displaced Workers

Relatively little is known about the effects of retraining and job search assistance for dislocated workers. If anything, this lack of knowledge is likely to become more important in the future as the public appears to be becoming increasingly sensitive to dislocations due to foreign

competition. Barnow and Gubits (pp. 71 and 94) advocate the use of random assignment to determine the effects of WIA dislocated worker training.

Our review of previous experiments indicates that retraining and job search assistance for dislocated workers has not been previously experimentally evaluated. There may be a good reason for this. In order to obtain an adequate sample size, the experiment would likely have to be conducted in geographic areas that have only recently experienced substantial dislocations and in which substantial numbers of the affected workers are interesting in retraining and job search assistance, presumably because they cannot readily find new jobs that they consider acceptable. In practice, this suggests that the experiment will have to be “on the drawing board” (i.e., designed in some detail well in advance of implementation) so that it can be very quickly fielded in relatively loose labor markets soon after major lay-offs occur.

An Experimental Evaluation of Job Placement for Disabled Workers

Barnow and Gubits state (p. 96) that “research is needed to determine if individuals with various types of disabilities are enrolled in [employment and training programs] in proportion to their need and/or prevalence in the population, whether the programs accommodate their disabilities adequately, and what the *impacts* (emphasis added) of these programs have been for individuals with disabilities.”³⁷ The British have an interesting program called the New Deal for Disabled Persons (NDDP) that might usefully be pilot tested in the United States and evaluated by random assignment. In NDDP, Job Brokers, which are a mixture of voluntary, public and private sector organizations, are contracted by the government to channel disabled persons into jobs. Although Job Brokers vary enormously in size, the geographic areas they cover, and in how they operate, most help clients with their job search, engage in job development, and attempt to increase client confidence in their ability to work. Many also attempt to develop client work skills and monitor client progress in jobs after they are placed, sometimes intervening when the client encounters problems on the job. Job Brokers receive an incentive reward from the government for each client they register, for each client they place in a job, and for each placed client who continues to work for at least six months.

Although random assignment could potentially be used to investigate the impacts of programs such as NDDP, it may be politically difficult to establish a control group because doing so would mean that some disabled persons who are actively seeking training would have to be denied these services at least temporarily. In fact, NDDP was initially to be evaluated by random assignment in the United Kingdom, but because of adverse publicity and political pressure, the idea was dropped. Consequently, the program is currently being evaluated by nonexperimental methods.

A UI Claimant Profiling Experiment

The Worker Profiling and Reemployment Services (WPRS) system uses profiling to predict which UI claimants are most likely to exhaust their benefit entitlement and should, therefore, receive mandatory reemployment services (e.g., structured job search, employment counseling, and retraining). To the best of our knowledge, WPRS has been tested through random assignment only in Kentucky.³⁸ The findings indicated that the WPRS system in Kentucky reduces weeks of UI receipt, increases earnings, and results in cost savings for the UI

Program. However, the evaluation also suggests that these findings mainly occur because the “threat” of having to participate in reemployment services causes some claimants to take jobs more quickly than they otherwise would, rather than receipt of the services themselves. Moreover, the treatment effects were not larger for claimants with higher profiling scores, suggesting the econometrically estimated profiling model used in Kentucky is not very efficient.

The Kentucky evaluation utilized a unique “tie-breaking” experimental design in which random assignment was possible because local UI offices in Kentucky face a capacity constraint on the number of claimants they can provide with reemployment services. Under these circumstances, only those claimants with relatively high profiling scores receive reemployment services. Under the “tie-breaking” experimental design utilized in Kentucky, if capacity was reached at a score of (say) 17 and 19 claimants had this score, but only 8 slots remained, then 8 claimants were randomly assigned to the treatment group and the remaining 11 claimants became controls. This “tie-breaking” design could be readily replicated in other states that administer their WPRS systems in a manner similar to Kentucky, and it should be. The Kentucky experiment was very inexpensive to conduct. Moreover, the profiling models and UI systems used in other states differ from Kentucky’s in various respects. It would be very useful to see if the findings for Kentucky occur elsewhere and if other econometric profiling models appear to be more efficient than Kentucky’s.

A Reemployment Account Fund Experiment

As discussed in section 2.0, the Bush administration has introduced legislation that would provide those UI claimants that profiling predicts are most likely to exhaust their benefit entitlements with reemployment account funds. According to the legislation, amounts in these funds could be used for three different purposes: (1) to purchase various reemployment services, (2) as income maintenance in those instances in which claimants exhaust their UI benefits (in other words, benefits would be provided beyond the standard 26 weeks, or beyond 39 weeks if the Extended Benefit Program is in effect), and (3) as reemployment bonuses in those cases in which claimants obtain employment within 13 weeks. The proposed program is expensive (an estimated \$3.6 billion over two years) and important aspects of it are untested. (As discussed in section 2.0, reemployment bonuses have been tested in four separate randomized experiments, but the findings were somewhat mixed.)

Before rolling out the proposed program nationally, it seems important to test it as a pilot program that is limited to just a few locations. Even if Congress decides to adopt the administration’s proposal, a strong case is seen for using random assignment trials to test some planned variations in the proposed scheme. The effects of the treatment tested in the pilot program could readily be evaluated experimentally by randomly assigning some UI claimants who are eligible to a program group and providing them with reemployment account funds and assigning the remainder of the eligible claimants to a control group. In fact, the evaluation could usefully go one step further by first establishing (say) three or four program groups and then randomly assigning eligible claimants among these program groups, as well as to a control group. The size of the reemployment account funds would vary among these program groups, allowing the fund size that produces the most favorable effects to be determined. The research sample should probably include UI claimants who would be ineligible for reemployment account

funds if the current version of the proposed legislation is passed, as well as those who would be eligible. As discussed below, this would allow a determination of whether the reemployment account funds have favorable effects on such individuals. If they do, the eligibility criteria could be suitably modified.

As mentioned above, even if reemployment accounts are established as national policy, we see strong arguments for using random assignment trials to measure the effectiveness of such accounts and to estimate the percentage of the UI caseload for which reemployment accounts are beneficial. Within a single state, for example, random assignment and planned program variation can be used to determine the effectiveness of reemployment accounts for samples of UI claimants who are “marginally eligible” for accounts under the state’s profiling formula. UI claimants could be classified in three groups—workers who are highly likely to exhaust benefits, workers who very unlikely to exhaust benefits, and workers who are in an intermediate category. Claimants in the last group could be randomly assigned to treatment eligibility status, allowing analysts to determine reliably the effectiveness of the treatment for UI claimants in this group. The proportion of UI claimants in the intermediate group could be systematically varied in different parts of the state to help officials identify the percentage of the UI caseload for which reemployment accounts are cost effective.

An Experimental Study of the Effects of Changing UI Benefits

The extent to which UI benefits increase the duration of unemployment and, hence, program costs has been studied for many years using nonexperimental methods, and there is strong agreement that the behavioral response to the program increases both the unemployment rate and program costs. According to a recent review of this research by Krueger and Meyer (p. 25),³⁹ the U.S. studies “imply an elasticity of duration with respect to the level of benefits in excess of 0.5.... [But] the elasticity estimates with respect to the potential duration (length) of benefits tend to be much lower.” Taken at face value, this suggests the possibility of reducing benefit levels by (say) 20 percent, but increasing the eligibility period so that UI duration rises by the same percent, as greater protection could then be provided the long-term unemployed at lower program costs. However, the elasticity estimates vary considerably from study to study. Thus, before major policy changes are based on such estimates, it would make sense to try to reduce the uncertainty about them.

This might be done through a social experiment. Random assignment could be used in a fairly straightforward fashion to determine the effects of varying the duration of UI Program eligibility and the weekly benefit amount. Bell (p. 153) suggests that this be done only by increasing the amount and duration of UI benefits in order to circumvent legal and ethical problems that would result if UI benefits were reduced for evaluation purposes. He argues that the findings should provide evidence on what would happen if UI benefits were reduced, as well as information on the effects of increasing benefits. The changes should be substantial enough to evoke a response that is sufficiently large to be statistically reliable. Thus, it might be less expensive to conduct the experiment in low-benefit states. Interpolation could be used to project responses to smaller changes than those tested.

In designing the experiment, it might be useful to randomly assign claimants to one of two or (possibly) three treatment groups, as well to a control group that would continue under existing program rules. Claimants assigned to the first treatment group would receive increased weekly benefit amounts; those assigned to the second group would be permitted to receive benefits for an extended period of time; and both the benefit amount and benefit duration would be increased for the third group. Comparisons of each of the first two groups with the control group would provide separate estimates of the effects of changing the benefit level and changing the benefit duration. The third treatment group would only be needed if an interaction effect results from changing both program parameters.

Of the three main financial determinants of UI generosity—the waiting period, the replacement rate, and the duration of weekly benefits—there has been little variation over time in the waiting period and the gross replacement rate. The waiting period for benefits has been either one or two weeks for several decades. The gross replacement rate, measured as the average weekly UI benefit divided by the average weekly total wage covered by UI, ranged from a low of 33.5 percent to a high of 37.5 percent between 1960 and 1997 with very little clear trend over the period. At the national level, the after-tax replacement rate has varied mainly because UI benefits were formerly tax-free and are now subject to taxation under the income tax. The duration of UI benefits has varied considerably more over time, mainly because extended benefits and Federal supplemental benefits have provided sizable extensions during recessions. Thus, for practical purposes, policymakers may have a larger stake in understanding the effects of benefit extensions, since this is the policy margin that is most frequently varied.

A Financial Incentive Experiment for Ex-Prisoners

In section 2.0, a small experiment was mentioned, the Living Insurance for Ex-Offenders (LIFE) Demonstration, which was conducted in the early 1970s. The tested program provided males who were recently released from prison and who returned to Baltimore with \$60 a week (in 1972 dollars) for 13 weeks, conditional on not being reimprisoned. Half of any earnings above \$40 were subtracted from the weekly payment and deferred to a later week, thereby slightly extending the payment period. The financial incentives were found to increase employment and decrease new arrests.⁴⁰

Given the current size of the prison population, it would appear useful to replicate this experiment in additional locations and with a larger sample. In doing this, the optimal payment amount and payment period could be investigated by randomly assigning the study sample to several different program groups, which would vary in terms of the size of the weekly payment amounts and the length of the payment period, as well as to a control group. A similar experiment that was targeted on persons who had recently undergone long-term drug rehabilitation could also be conducted.

A Random Assignment Evaluation of the Federal Bonding Program

Although the Federal Bonding Program, which is briefly described in section 2.0, has existed since 1966, little is known about its effects. However, it is apparent that it is not well-known and that it is little used. Barnow and Gubits (p. 86) suggest that the program be evaluated. Such an evaluation could be accomplished by randomly assigning some eligible individuals to a

control group that is made temporarily ineligible for the bond or, perhaps more realistically, by using random assignment to vary the bond amount among individuals who qualify for the program. The length of time the bond remains effective could be similarly varied.

Given the previous lack of utilization of the program it seems likely that such an experiment would have difficulty in enrolling a sample of sufficient size to obtain statistically reliable findings. Thus, the first step in conducting the experiment should be to publicize the program to the eligible “at-risk” population of job applicants (i.e., ex-offenders, those dishonorably discharged from the military, recovering drug and alcohol abusers, and low income persons without a work record) and encourage them to take advantage of it. The U.S. Employment Service could, perhaps, be enlisted to do this. In addition, prisoners could be informed about the Federal Bonding Program as they are being released. The applicant would then tell perspective employers about the availability of the bond. If an adequate number of at-risk job applicants could not be enlisted, the experiment could simply be aborted; even then, however, very useful information about the value of the program will have been obtained.

Even if the sample is of sufficient size, it may still turn out that the program is ineffective. For example, a previous experiment in which disadvantaged job applicants informed employers that they would receive tax credits or direct cash payments if they hired the worker was found to have a negative effect on employer hiring decisions, possibly because these demand side incentives stigmatized the very persons they were intended to help.⁴¹ Of course, the object of the Federal Bonding Program is to reduce stigma that would otherwise exist—that is, it is intended to reduce the risk that employers perceive they incur by hiring certain individuals. However, it is not known whether, in practice, it actually has this helpful effect. One purpose of the experiment would be to determine whether the program reduces or increases stigma.

A U.S. Replication of the Canadian SSP Experiment

The Canadian Self-Sufficiency Project (SSP) was a recently completed random assignment experiment in New Brunswick and British Columbia and targeted on single-parent welfare recipients. SSP tested a temporary monthly cash supplement equaling half the difference between a participant’s earnings and an annual earnings benchmark of C\$30,000 in New Brunswick and C\$37,000 in British Columbia if they worked at least 30 hours a week. The supplement was available only to single parents who had been on welfare for at least a year and had to be taken up within a year of being offered. Once taken up, it was available for a maximum of three years. The tested program had substantial positive effects on full-time employment, earnings, and family income. Although these effects tended to disappear after eligibility for the program ended for participants who were only offered financial incentives, they continued for a smaller second group that was offered employment-related services (job clubs, self-esteem workshops, and others) in addition to financial incentives. The costs to the government of SSP were well under the increases in income enjoyed by participants. Thus, the program seemed to be an efficient transfer mechanism.

Given the very positive effects of the SSP program, it is probably worth replicating it in the United States. If it is replicated however, several program groups, as well as a control group, should be established and key program parameters (i.e., the earnings benchmark, the length of

time on welfare before program eligibility begins, the length of the period between the offer and the required take-up, and the length of time benefits continue) should be allowed to vary among the program groups.⁴² Because SSP only tested a single program, it is not known whether the impacts are sensitive to key program parameters. Hence, it would be useful to vary these parameters in a United States replication, as well as test the Canadian version of the program. The Canadian results imply that for purposes of the replication the financial incentives should be combined with services. One technical problem in conducting the experiments would be meshing the financial incentives with the Federal EITC and, in some cases, state EITCs. One way of doing this is simply to make the benefits received under EITC independent of SSP payments, and vice versa.

4.2 Lower Priority Experiments

1. Barnow and Gubits (p. 42) recommend that further research be conducted to determine the effectiveness of alternative training service sequencing strategies. A head-to-head comparison along the lines of the approach outlined in item 2 above could be used to do this. However, in item 2, we suggest that the approach be used to compare alternative training packages. We do not consider a comparison of alternative sequencing strategies as important.
2. Barnow and Gubits (p. 61) and many others have mentioned that a possible topic for evaluation is post-training impact decay rates and how they vary across services and client types. This topic is important for cost-benefit assessments of training programs. If short followup periods are used to evaluate training programs, which is typically the case, extrapolating program impacts beyond the followup period must be based on educated guesses. Unfortunately, policy decisions that are made on the basis of short followup periods and bad guesses might be inappropriate. However, greater knowledge about decay rates does not require additional evaluations, but the collection of longer followup data in the evaluations that are conducted. Obtaining such data could be made part of the ongoing WIA experiment suggested in item 1.
3. Customized training programs are tailored to the particular needs of specific employers, but wholly or partially funded by government agencies. Most states operate such programs and they are also permissible under WIA. Employers help develop the program curriculum and recruit and select the potential trainees. Indeed, many of the workers trained by the state programs are already employed by the firms for which the training is being provided. Barnow and Gubits review the literature on these programs (pp. 53-56) and conclude that while they appear to have considerable promise, little is known about their impacts. Thus, they suggest that “if possible, an impact evaluation should be conducted on customized training programs to determine if they are as good as they appear to some observers....” (p. 56). They do not suggest, however, that the impact analysis be based on randomized assignment. Unfortunately, such an analysis is probably infeasible. For example, one approach to such an experiment would involve two steps: employers would first recruit and select workers who are eligible for customized training; and then a random subset of these workers would actually receive the training, with the remaining workers constituting a control group. It seems unlikely that many employers would be willing to cooperate with conducting such an experiment, which would involve recruiting substantially more workers

for training than are needed. A second approach would be to randomly assign some employers who are interested in undertaking customized training to a program group for whom the training is provided and the remaining employers to a control group for whom it is not provided. Doing this would probably require special legislation, which seems unlikely to be acceptable. Moreover, the number of employers who are involved in customized training is relatively small, suggesting that obtaining a sample of sufficient size could be difficult.

4. In recent years, Community-Based Organizations (CBOs) have been used to provide various services to government-operated programs, including job development, job placement, and employment and training. In providing these services, they have often been funded by contracts with local government agencies or by a DOL grant. Because government agencies also deliver the same services, it would be useful to know which delivery mechanism is more effective. An examination of this topic could be done by randomly assigning individuals needing such services to CBOs and government agencies. Such an experiment would have to be limited, however, to geographic areas where both types of organizations are delivering the same service.⁴³ In evaluating the experiment, differences between the two types of organizations in costs, as well as benefits, should be estimated.
5. Over the last two decades, youths with delinquency or substance abuse problems have sometimes been sent to so-called “boot camps.” Because of the severe discipline that boot camps impose, they have been highly controversial. To the best of these authors’ knowledge, however, they have never been subject to an experimental evaluation. Presumably, such an experiment could be readily conducted by simply randomly assigning the relevant target population to either a boot camp or to more conventional treatment facilities (e.g., a local drug treatment facility).⁴⁴ However, it may be difficult to collect reliable data on some of the outcomes of interest, which presumably would include criminal activity and drug use, as well more conventional outcomes such as employment, earnings, and transfer program receipts.
6. Of all the Welfare-to-Work Programs evaluated by random assignment in the United States, two stand out as having exceptionally large effects: one in Riverside, California; and the other in Portland, Oregon. Key features of the Riverside Program were replicated in Los Angeles and the resulting program was evaluated by random assignment to determine if Riverside’s apparent success was due to factors that were unique to that site (e.g., client or labor market characteristics or especially strong administrative leadership) or could occur in a very different setting. Although the effects found in Los Angeles were not as dramatic as those measured in Portland, they were quite positive. Barnow and Gubits (pp. 75 and 95) suggest that it might be useful to also replicate the Portland Welfare-to-Work Experiment in a different setting. This should be feasible as many of the key features of the Portland Program are transferable. For example, participants in the Portland Program were told that the goal for them was to obtain a job, but that they should wait until they are able to take a good job. In keeping with this philosophy, clients who were in need of more skills were encouraged to enroll in education and training initially and look for a job later.
7. In their report, the DOL Expert Panel states that research should be conducted to determine “what transportation strategies are most effective in assisting individuals who lack reliable and affordable transportation to find and retain employment” (p. 176). This is an important

issue in some communities because existing bus lines do not operate directly between areas where low income workers live (e.g., central cities) and areas where job openings exist (the suburbs) and relatively few central city residents own automobiles. Although it may sometimes be possible to make the trip by transferring among buses, the time and effort required may discourage many workers from seeking or accepting jobs in the suburbs. The major potential alternatives available to the government to address this problem is the provision of additional bus routes, the provision of van pools, and the subsidization of the purchase and maintenance of private automobiles. Unfortunately, only an evaluation of the third of these alternatives is very amenable to the random assignment of individuals. One can readily imagine randomly assigning interested individuals to a program group that receives help in obtaining and maintaining an automobile and a control group that does not; but it is difficult to imagine allowing some individuals access to enhanced bus or van service and not allowing their neighbors the same access. Nonetheless, the first and second alternatives could potentially be evaluated by randomly assigning entire neighborhoods and providing “program neighborhoods” with improved bus or van service and not doing the same for “control neighborhoods.” To obtain statistically reliable findings, however, a substantial number of neighborhoods would have to be randomly assigned. Doing this may not be very feasible, either practically or politically. Thus, it seems unlikely that an elaborate experiment of this sort would be worth undertaking.

8. The DOL Expert Panel (p. 175) recommends that research be conducted on the usefulness of transitional employment in conjunction with other services in helping welfare recipients obtain higher wages and permanent employment. As discussed in section 2.0, experimental evidence already exists that subsidized jobs and work experience jobs, which are one form of transitional employment, are sometimes effective. Moreover, welfare recipients have been targeted by some of the programs evaluated by these experiments. Jobs with temporary help agencies provide another important type of transitional employment. In principle, it would probably be useful to conduct an experiment that investigated the longer-term effects of employment through such agencies. However, it would likely prove difficult to field such an experiment because the cooperation of temporary help agencies that had more applicants than open positions would be needed. Such agencies would have to agree to use randomization in assigning applicants to job openings.
9. As mentioned in section 2.0, it is possible in principle to assign the One-Stop Career Centers randomly to different treatment groups and vary these groups in terms of staff size and quality, software and equipment, or services provided to determine what combination is most effective. However, such an experiment may not be very practical. Obtaining reliable findings would require randomly assigning a substantial number of One-Stop Career Centers. Assigning different treatments to a sizeable number of different One-Stop Career Centers and ensuring that they actually adhere to the treatment they are assigned would be extremely difficult. This would be especially true if the One-Stop Career Centers in the study sample cross state boundaries, as the cooperation of individual state governments would then be required. Moreover, because many users of One-Stop Career Centers mainly rely on the computerized information they can obtain there, it is not obvious that sufficient identifying information on users can be obtained to obtain the follow-up data (e.g., on earnings) needed to conduct an evaluation.

10. Barnow and Gubits (p. 91) suggest that an impact evaluation of the Indian and Native American Program (a training program funded under the WIA) be considered, although they do not specifically recommend using random assignment do to this. As discussed in section 2.0, a social experiment is probably technically feasible. Indeed, in the absence of random assignment, it is not obvious how a viable comparison group could be obtained. However, most Indian reservations have small populations. Therefore, it would probably be difficult to obtain an adequate sample without conducting the experiment in a substantial number of different reservations. Unfortunately, closely monitoring a social experiment on a large number of often geographically isolated Indian reservations to prevent cross-over and control contamination may require more resources than are warranted by the value of the findings that would be obtained.
11. In their report (p. 174), the DOL Expert Panel indicated that a high priority should be placed on determining the effectiveness of post-employment services and financial incentives on keeping workers with unstable work patterns employed. These authors agree and believe that randomized social experiments should be used to do this. However, as mentioned in section 2.0, random assignment is, in fact, already being used to test the effects of post-employment services and financial incentives on job retention in both the United States and in the United Kingdom. Additional random assignment experiments that examine this topic should probably wait until findings from the current experiments become available.

4.3 Postscript

Unemployed, out-of-school youth are clearly an important target group. Unfortunately, very few programs appear effective in increasing the employment or earnings of these youngsters. However, any experiments that would test programs that are targeted on this group are not suggested because we do not know of innovative interventions that are likely to be successful. To obtain such ideas, DOL should solicit proposals from municipalities, colleges, and entrepreneurs on specific employment- or training-orientated policies that might enhance the labor market performance of out-of-school youth. DOL funds could then be granted to agencies or organizations that write winning proposals with the stipulation that the proposed treatment be tested using a rigorous evaluation methodology. In practice, this will probably mean that the proposed policy would be evaluated using random assignment among other techniques. Apparently, successful interventions could be replicated elsewhere, again using random assignment, to see if their success can be duplicated in other settings.

Endnotes

¹ An “internally valid” estimate is an unbiased measure of the treatment effect in the sample actually enrolled in an experiment. An “externally valid” estimate is a treatment-effect estimate that can be validly extrapolated to the entire population represented by the sample enrolled in the experiment. Some experimental estimates may not be internally valid, possibly because treatment has not been assigned randomly or because attrition produced treatment-group and control samples that are not comparable (see below). In addition, some internally valid estimates may lack external validity. One reason is that a treatment offered to a small experimental sample may not correspond to any treatment that could actually be provided to a broad cross-section of the population. The pros and cons of experimentation are debated by Gary Burtless, “The Case for Randomized Field Trials in Economic and Policy Research,” *Journal of Economic Perspectives*, 9, 2 (Spring 1995) pp. 63-84, and James J. Heckman and Jeffrey A. Smith, “Assessing the Case for Social Experiments,” *Journal of Economic Perspectives*, 9, 2 (Spring 1995) pp. 85-110.

² See Charles Michalopoulos, David Card, Lisa A. Gennetian, Kristen Harknett, and Philip K. Robins, *The Self-Sufficiency Project at 36 Months: Effects of a Financial Work Incentive on Employment and Income* (Ottawa, Canada: Social Research and Demonstration Corporation, 2000). The British Columbia target wage in 1994 was \$37,000 (Canadian), which is equivalent to \$44,400 (Canadian) in 2003. The Organization for Economic Cooperation and Development’s (OECD) estimates of PPP exchange rates was used to convert 2003 Canadian dollars into equivalent 2003 U.S. dollars.

³ See Larry L. Orr, Howard Bloom, Stephen Bell, Fred Doolittle, Winston Lin, and George Cave, *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study* (Washington, DC: Urban Institute, 1996).

⁴ James J. Heckman and Alan B. Krueger, *Inequality in America: What Role for Human Capital Policies?* (Cambridge, MA: MIT Press, 2003).

⁵ Virginia Knox, Cynthia Miller, and Lisa A. Gennetian, *Reforming Welfare and Rewarding Work: Final Report on the Minnesota Family Investment Program* (New York, NY: Manpower Demonstration Research Corporation, 2000).

⁶ Orley Ashenfelter and David Card, “Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs,” *Review of Economics and Statistics*, 67, 4 (November 1985) pp. 648-660, and Burt Barnow, “The Impact of CETA Programs on Earnings: A Review of the Literature,” *Journal of Human Resources*, 22, 2 (Spring 1987) pp. 157-193. After these articles were written, a number of researchers proposed nonexperimental estimators to reduce or eliminate some of the shortcomings of older estimators. None, however, have proven successful in consistently eliminating the discrepancies between an experimentally based estimate and the nonexperimental estimate of the same treatment effect. For a review of this literature and an examination of 12 case studies intended to replicate treatment effect estimates of workforce interventions, see Steven Glazerman, Dan M. Levy, and David Myers, “Assessing Alternatives to Social Experiments,” *Annals of the American Academy of Political and Social Science*, 589 (September 2003) pp. 63-93.

⁷ See Gary Burtless and Larry L. Orr, “Are Classical Experiments Needed for Manpower Policy?” *Journal of Human Resource*, 21, 4 (Fall 1986) esp. pp. 621-624, and the essays in Alice M. Rivlin and T. Michael Timpane, editors, *Ethical and Legal Issues of Social Experimentation* (Washington, DC: The Brookings Institution, 1975).

⁸ One potential within-sample error that is peculiar to demonstrations including experiments is Hawthorne effects, the possibility that the behavior of participants in a demonstration could be influenced by the knowledge that they are part of a trial. In an experimental setting, Hawthorne effects will cancel out if program and control groups are affected similarly. However, there is little information available about whether Hawthorne effects bias findings from experiments. In the one experiment in which there was an attempt to measure Hawthorne effects by informing one control group that they were part of a demonstration and not informing a second control group that was otherwise treated similarly, no Hawthorne effect was detected. See Daniel H. Klepinger, Terry R. Johnson, Jutta M. Hoesch, and Jacob M. Benus, *Evaluation of the Maryland Unemployment Insurance Work Search Demonstration*, Battelle Memorial Institute. Perhaps even more significantly, social scientists now doubt whether a “Hawthorne effect” actually occurred at the Western Electric Hawthorne Works plant in Cicero, Illinois, in the famous 1927–1932 “experiment” in which the phenomenon was first observed. Using modern statistical methods in a careful reanalysis of the surviving data, Stephen R.G. Jones finds little or no evidence for any such effect. See S.R.G. Jones, “Was There a Hawthorne Effect?” *American Journal of Sociology*, 98, 3 (November 1992) pp. 451-468.

⁹ Robert Moffitt, “The Effect of Employment and Training Programs on Entry and Exit from the Welfare Caseload,” *Journal of Policy and Management*, 15 (1996) pp. 32-50, and others have argued that program entry and deterrent effects could be substantial. However, findings from nonexperimental attempts to measure these effects, which have generally relied on aggregate-level time-series studies of programme applications, are mixed and inconclusive (for a review of these studies, see Daniel D. Friedlander, David H. Greenberg, and Philip K. Robbins, “Evaluating Government Training Programs for the Economically Disadvantaged,” *Journal of Economic Literature*, 35 (1997) pp. 1809-1855. In the only attempt to use experimental methods to measure entry effects (G. Berlin, W. Bancroft, D. Card, W. Lin, and P. K. Robins, *Do Work Incentives have Unintended Consequences? Measuring ‘Entry Effects’ in the Self-Sufficiency Project* (Ottawa, Canada: Social Research Demonstration Corporation, 1998)), Canadian single parents who had been newly added to the welfare rolls and also randomly allocated to a treatment group were told that if they remained on welfare for the next 12 months, they would subsequently qualify for very generous earnings supplements provided they then worked full time. The control group was not given this information, as they were not eligible for the earnings supplement. After the 12 months, 3.1 percent more of the treatment group than the control group were still on the welfare rolls.

¹⁰ The Job Corps Experiment had an important advantage in comparison with the JTPA Experiment, however. It enrolled a stratified random sample of applicants from a national applicant pool, and thus its estimates of program impact reflect responses to the average program throughout the nation. In contrast, the JTPA Experiment enrolled treatment and control group applicants in only 16 local areas around the United States that were willing to participate. The findings of the JTPA Experiment provide a reliable estimate of the average impact of program services in those sites, but they provide a less reliable guide to the average JTPA impact throughout the entire United States.

¹¹ Gary Burtless and David Greenberg, “Inferences Concerning Labor Supply Behavior Based on Limited-Duration Experiments,” *American Economic Review*, 72, 3 (June 1982) pp. 488-497.

¹² In 1961, for example, the U.S. Social Security Program was reformed to permit men between 62 and 64 to claim reduced, early retirement pensions. The incentive for men to retire between ages 62 and 64 was instantaneously changed to encourage additional retirements at those ages. Yet it took several years before there was a large increase in the proportion of men who left the labor force at age 62. By 1980, age 62 was the most popular age for leaving the labor force, far more popular than retirement at age 65. Using nonexperimental econometric methods, it is impossible to determine reliably how much of the change in retirement behavior is due to the direct effects of altered social security incentives and how much is due to shifts in retirement preferences. Thus, nonexperimental research methods may have few advantages over randomized trials in uncovering the effects of policy-induced changes in preferences or social norms.

¹³ Burtless and Orr (op. cit.) refer to this inference problem as “queuing bias.” A small-scale experiment adds only a small number of people to the queue of job seekers for a limited number of jobs. Communitywide implementation of the same treatment would add a much larger number of people to the queue, and the treatment might therefore have a smaller beneficial effect on each person in the queue.

¹⁴ Jeremy Lise, Shannon Seitz, and Jeffrey Smith try to combine the experimental findings from the Canadian Self-Sufficiency Project (SSP) with a nonexperimental general equilibrium model to infer the full effects of implementing the SSP model on the Canadian labor market. See Lise, Seitz, and Smith, “Equilibrium Policy Experiments and the Evaluation of Social Programs,” (Working Paper 10283, National Bureau of Economic Research, January 2004).

¹⁵ An experiment could shed light on the general equilibrium effect of a bonus, but the experiment would be very costly. Experimenters could offer bonuses of varying generosity to unemployed workers in different communities. (The bonuses would have the same value within a community.) Researchers would then compare average unemployment durations and starting wage offers across the communities to determine the general equilibrium effects of different bonus offers.

¹⁶ A more detailed discussion than we have space for here can be found in David Greenberg, Donna Links, and Marvin Mandell, *Social Experimentation and Public Policymaking* (Washington, DC: Urban Institute, 2003).

¹⁷ Karl Ashworth, Andreas Cebulla, David Greenberg, and Robert Walker, “Meta-evaluation: Discovering What Works Best in Welfare Provision,” *Evaluation*, 2004.

¹⁸ David H. Greenberg, “Conceptual Issues in Cost-Benefit Analysis of Welfare-to-Work Programs,” *Contemporary Policy Issues*, 10 (1992) pp. 51-64.

¹⁹ David Greenberg, Donna Links, and Marvin Mandell, *Social Experimentation and Public Policymaking*, (Washington, DC: Urban Institute Press, 2003).

²⁰ Eligibility for Project GATE is open to virtually all adults. Thus, it has multiple target groups and, consequently, is counted four times in table 1. In addition, one of the other experiments had two target groups and thus was counted twice in table 1. There have been only three experimental evaluations of entrepreneurship and self-employment programs.

²¹ However, as discussed below, the Bush administration has introduced legislation to provide certain UI claimants with personal reemployment accounts, which, under certain circumstances, can be received as reemployment bonuses. It is not clear whether the reemployment bonus experiments influenced this initiative.

²² However, there is evidence that from a broader social perspective, as opposed to the narrow perspective of the UI system, reemployment bonuses can be cost-beneficial (see Philip K. Robins and Robert G. Spiegelman, editors, *Reemployment Bonuses in the Unemployment Insurance System: Evidence from Three Field Experiments*, (Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2001).

²³ One would expect the measured effects on earnings to be small as the following illustration suggests. Imagine that a tested treatment increases high school graduation by 10 percentage points and those who graduate who otherwise would have dropped out enjoy an earnings increase of 20 percent. The measured impact on earnings for the entire program group would be only 2 percent (.1 x .2).

²⁴ David Greenberg and Mark Shroder, *Digest of Social Experiments, Third Edition* (Washington, DC: Urban Institute Press, 2004).

²⁵ The preliminary results turned out to be final. Because of funding constraints, DOL cancelled the experiment in May 1981. Analysis of the experiment was therefore based on information collected from the management information system rather than from administrative records or interviews, as originally planned. See Gary Burtless, "Are Targeted Wage Subsidies Harmful? Evidence from a Wage Voucher Experiment," *Industrial and Labor Relations Review*, 39, 1 (October 1985) pp. 105-114.

²⁶ See Sheena McConnell and Steven Glazerman, *National Job Corps Study: The Benefits and Costs of Job Corps*, (Princeton, NJ: Mathematica Policy Research, Inc., 2001), Figure III.1, p. 44, and surrounding discussion.

²⁷ The uncertainty about the longer term impact of the Job Corps Program led James Heckman and Alan Krueger to reach nearly opposite conclusions about the value of Job Corps services to young people enrolled in the treatment group of the Job Corps Experiment. See Heckman and Krueger (op. cit.), esp. pp. 43-48 and 190-94.

²⁸ See Greenberg, Links, and Mandell (op. cit.), Report 7.

²⁹ Christopher J. O'Leary, Paul Decker, and Stephen A. Wandner, "Reemployment Bonuses and Profiling," (Working Paper 98-51, W.E. Upjohn Institute for Employment Research, 1997).

³⁰ For a review and synthesis of the literature on the impact of welfare reform, declining unemployment, and the EITC on caseload decline and other trends in the population served by AFDC and TANF, see Rebecca M. Blank, "Evaluating Welfare Reform in the United States," *Journal of Economic Literature*, 40, 4 (December 2002) pp. 1105-1166, and Jeffrey Grogger, Lynn A. Karoly, and Jacob A. Klerman, *Consequences of Welfare Reform: A Research Synthesis* (Santa Monica, CA: The RAND Corporation, 2002).

http://www.acf.hhs.gov/programs/opre/welfare_reform/reform_cover.html

³¹ Burt S. Barnow and Daniel B. Gubits, section III, "Review of Recent Pilot, Demonstration, Research, and Evaluation Initiatives to Assist in the Implementation of Programs Under the Workforce Investment Act," in *Five-Year Strategic Research Plan for Pilots, Demonstrations, Research, and Evaluations (July 2002-June 2007)*, (Washington, DC: U.S. Department of Labor, forthcoming).

³² Stephen Bell, section IV, "Review of Alternative Methodologies for Employment and Training Research" in *Five-Year Strategic Research Plan for Pilots, Demonstrations, Research, and Evaluations (July 2002-June 2007)*, (Washington, DC: U.S. Department of Labor, forthcoming); and DOL Expert Panel, sections V and VI, "High Priority Topics/Subject Areas for Pilots, Demonstrations, Research, and Evaluations" in *Five-Year Strategic Research Plan for Pilots, Demonstrations, Research, and Evaluations (July 2002-June 2007)*, (Washington, DC: U.S. Department of Labor, forthcoming).

³³ Larry L. Orr, Howard Bloom, Stephen Bell, Fred Doolittle, Winston Lin, and George Cave, *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study* (Washington, DC: Urban Institute, 1996).

³⁴ For an illustrative example of the usefulness of multilevel modeling in the context of welfare-to-work evaluations, see Howard S. Bloom, Carolyn J. Hill, and James Riccio, "Linking Program Implementation and Effectiveness: Lessons from a Pooled Sample of Welfare-to-Work Experiments," *Journal of Policy Analysis and Management*, 22, 4 (2003) pp. 551-575.

³⁵ For a detailed discussion, see Burt S. Barnow, “The Effects of Performance Standards on State and Local Programs,” in *Evaluating Welfare and Training Programs*, edited by Charles F. Manski and Irwin Garfinkel (Cambridge, MA: Harvard University Press, 1992) pp. 277-309.

³⁶ We are indebted to James Riccio for suggesting conducting a performance standards experiment.

³⁷ As mentioned in section 1.2, an experimental evaluation was recently completed of Project Network, a program that provided training among other services. However, only a relatively small subset of Project Network participants actually received training. Thus, the experiment was not an evaluation of training for the disabled.

³⁸ Dan Black, Jeffrey Smith, Mark Berger, and Brett Noel, “Is the Threat of Reemployment Services More Effective than the Services Themselves? Evidence from Random Assignment in the UI System,” *American Economic Review*, 93, 4 (September 2003) pp. 1313-1327.

³⁹ Alan B. Krueger and Bruce D. Meyer, “Labor Supply Effects of Social Insurance,” (Working Paper 9014, National Bureau of Economic Research, 2002).

⁴⁰ Charles D. Maller and Craig Thornton, “Transitional Aid for Released Prisoners: Evidence from the LIFE Experiment,” *Journal of Human Resources*, 13, 2 (Spring 1978) pp. 208-236.

⁴¹ Gary Burtless, “Are Targeted Wage Subsidies Harmful? Evidence from a Wage Voucher Experiment,” *Industrial and Labor Relations Review*, 39 (October 1985) pp. 105-114.

⁴² We are indebted to Philip Robins for suggesting the SSP replication experiment to us.

⁴³ The only previous experiment that we know of that examined the relative effectiveness of alternative service delivery organizations is the Public Versus Private Sector Jobs Demonstration Project. In this experiment, which was conducted during the late 1970s in five sites located throughout the U.S., youths were randomly assigned to public sector employers or to private sector employers for 25 weeks of work experience. Although they were fully subsidized for the 25 weeks, the key outcome that was examined was their subsequent employment. Unfortunately, the experimental data were subject to severe attrition bias. (See James F. Gilsinan, “Information and Knowledge Development Potential,” *Evaluation Review* 8, 3 (June 1984) pp. 371-388.) Because employer wage records are much more readily available today than they were in the 1970s, this bias could be minimized in any future experiment.

⁴⁴ We are indebted to John Wallace for suggesting both this experiment and the preceding one.

Appendix A: Topologies and List of Experiments

Table 1. Topologies Used for Classification of Experiments

Program Areas	
DOL Training and Education Programs	A
Welfare-to-Work Programs	B
Other Non-DOL Training and Education Programs	C
Subsidized Employment and Work Experience Programs	D
Trade Adjustment Assistance	E
Entrepreneurship and Self-Employment Programs	F
Unemployment Insurance	G
Labor Exchange Programs	H
Job Search and Placement Programs	I
Minimum and Living Wage Programs	J
Demand Side Subsidies	K
Supply Side Subsidies	L
Loans and Subsidies for Education and Training	M
Federal Bonding Program	N

Target Groups	
Unemployed Adults <i>(excluding experiments targeted only at welfare recipients)</i>	1
Unemployed, Out-of-School Youth <i>(excluding experiments targeted only at welfare recipients)</i>	2
Welfare Recipients	3
Non-workers with Severe Employment Barriers not Elsewhere Classified <i>(includes disabled, mentally impaired, homeless, ex-convicts, recovering substance abusers)</i>	4
In-School Youth	5
Currently Employed People Who Earn Low Wages <i>(includes migrant and seasonal workers, older workers, and unemployed workers and former UI claimants and AFDC recipients who have found jobs)</i>	6
Native Americans	7

Table 2. List of Experiments, Descriptions, and Classifications

Experiment	Description	Target group	Program area	Other target groups	Other program area
National Job Training Partnership Act (JTPA) Study	This demonstration, conducted from 1987 to 1991 in 16 sites located throughout the U.S., tested the effects of the JTPA Title II Program's employment and training services on a large sample of economically disadvantaged adults and youths.	1	A	2	-
General Education in Manpower Training	This demonstration, conducted from 1964 to 1966 in Norfolk, Virginia, tested the effects of technical and general education in a medium-sized sample of unemployed males.	1	A	-	-
Buffalo Dislocated Worker Demonstration Program	This demonstration, conducted from 1982 to 1984, tested the effects of employment services (job search assistance, training, on-the-job training) on a medium-sized sample of laid-off workers.	1	A	-	-
Nevada Concentrated Employment Program (CEP)	This demonstration, conducted from 1988 to 1989, tested the effects of enhanced employment services and training on a large sample of unemployed workers.	1	A	-	-
Job Search Assistance Demonstration	This demonstration, conducted between 1995 and 1996 in the District of Columbia and Florida, tested the effects of three reemployment service packages on a large sample of Unemployment Insurance (UI) claimants	1	A	-	I
Chicago Housing Voucher Lottery	This ongoing experiment, which began in 1997, is testing the effects of Section 8 rental subsidies (housing vouchers).	1	C	6	-
Texas Worker Adjustment Demonstration	This demonstration, conducted from 1984 to 1985, tested the effects of assisted job search, job clubs, and training on a large sample of unemployed/displaced workers.	1	C	-	-
Jobs-Plus	Jobs-Plus is an ongoing quasi-experiment (public housing projects were randomly assigned), which began in 1998. Treatment projects receive a three-pronged intervention strategy: (1) saturation availability of employment support activities, including job search assistance, child care, transportation, some education and training, (2) financial incentives, which tend to break the connection between earnings and public housing rent, and (3) activities to strengthen social capital, including support groups and child care cooperatives.	1	C	-	L

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Project GATE (Growing America through Entrepreneurship)	This ongoing demonstration, which is currently being conducted at seven sites, is testing the effects of providing classroom training and one-on-one technical assistance, as well as help in applying for loans, to persons interested in developing or expanding their own businesses. The program is open to anyone over 18 years of age who attend an orientation session.	1	F	3,4,6	-
Project HOPE (Head Start Opportunities for Parents through Employment)	This demonstration, conducted between 1989 and 1991 in Columbus, Ohio, tested the effects of intensive case management, life-skills and job-readiness training on a small sample of parents with children enrolled in Head Start. Serious failures occurred in attempts to obtain followup information from the sample.	1	C	-	-
Washington and Massachusetts Unemployment Insurance Self-Employment Work Search Demonstrations	This demonstration, conducted from 1989 to 1991, tested the effects of business start up services, including financial assistance, counseling, and workshop sessions on a large sample of the recently unemployed.	1	F	-	-
Claimant Placement and Work Test Demonstration	This demonstration, conducted in 1983 in Charleston, South Carolina, tested the effects of a deadline to register for work with the Employment Service, mandatory supervised job search, and job placement services on a large sample of UI claimants.	1	G	-	-
Illinois Unemployment Insurance Incentive Experiment	This demonstration, conducted from 1984 to 1985, tested the effects of a reemployment bonus incentive on a large sample of UI clients.	1	G	-	-
Pennsylvania Reemployment Bonus Demonstration	This demonstration, conducted in 1988 and 1989, tested the effects of a reemployment bonus payment on a large sample of UI recipients.	1	G	-	-
Washington State Reemployment Bonus Experiment	This demonstration, conducted from 1988 through 1990, tested the effects of a reemployment bonus on a large sample of the unemployed who were applying for UI benefits.	1	G	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Washington State Alternative Work Search Experiment	This demonstration, conducted from 1986 through 1988, tested the effects of alternative work-search policies on a large sample of UI claimants.	1	G	-	I
Maryland Unemployment Insurance Work Search Demonstration	This demonstration, conducted from 1994 to 1996, tested the effect of four alternative work search policy requirements on a large sample of new UI claimants.	1	G	-	I
Delaware Dislocated Worker Pilot Program	This demonstration, conducted in 1983, tested the effects of job search assistance and employment counseling on a medium-sized sample of UI claimants.	1	G	-	I
New Jersey Unemployment Insurance Reemployment Demonstration	This demonstration, conducted from 1986 to 1987, tested the effects of job search assistance, training, and a reemployment bonus payment on a large sample of dislocated (unemployed) workers.	1	G	-	A
Reemploy Minnesota (REM)	This demonstration, conducted from 1988 to 1990, tested the effects of intensive case management services on a large sample of unemployed workers.	1	G	-	I
Kentucky Worker Profiling and Reemployment Services (WPRS) Experiment	This evaluation, conducted from 1994 to 1999, used a unique “tie-breaking” experimental design to estimate the impacts of the WPRS system on a large sample of UI claimants. Subjects were followed for 2 years.	1	G	-	-
U.S. Employment Service (ES) Effectiveness of Counseling Pilot Study	This demonstration, conducted from 1975 to 1976 in Minneapolis, Salt Lake City, and West Palm Beach, tested the effects of employment counseling on a medium-sized sample of ES clients.	1	H	-	-
Carbondale (IL) Job-Finding Club	This demonstration, conducted in 1973, tested the effects of supervised group job search assistance on a small sample of the unemployed.	1	I	-	-
Bridges to Work	In this ongoing experiment, which began in 1996, work-ready, low-income, inner-city residents can receive the following services—job placement, transportation, and supportive services such as child care and counseling—in order to place participants in suburban jobs.	1	I	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Minority Male Opportunity and Responsibility Program	This demonstration, conducted between 1991 and 1994 in Milwaukee, tested the effects of intensive case management on a small sample of unemployed minority males. The design was not implemented as planned, owing to a number of difficulties.	1	I	-	-
Individual Training Accounts (ITAs)	This ongoing experiment, which began in 2001, is testing three different approaches to ITAs with respect to resources expended on unemployed adults and dislocated workers in six sites across the United States: (1) structured trainee choice in which counselors play a central role directing trainees to specific high return training costing up to \$8,000, (2) guided trainee choice in which the counselor will play a less central role with a fixed training account of up to \$3,000, and (3) wide trainee choice, essentially a voucher program of up to \$3,000. Trainees are free to spend allotted resources on any state-approved training program and on related expenses approved by local counselors.	1	M	-	-
Seattle-Denver Income Maintenance Experiment	This demonstration, conducted from 1970 to 1977, tested the effects of both a negative income tax (at various levels of income guarantee and tax rate) and subsidized vocational counseling and training on a large sample of low-income families. Subjects were followed for up to 5 years. It was designed so that the impacts of the subsidized counseling and training could be measured separately from the impacts of the negative income tax.	1	M	6	
New Chance	This demonstration, conducted from 1989 to 1992 in 16 sites in 10 states, tested the effects of a comprehensive program emphasizing both human capital and personal development services on a large sample of young mothers on welfare who were high school dropouts and on their children.	2	A	3	-
Success Connection	This demonstration, conducted from 1991 through 1993 in Yakima Valley, Washington, tested the effects of adventure-based counseling and exposure to college on a small sample of at-risk rural youth.	2	A	-	-
National Job Corps Study	This study, conducted from 1994 to 1996 at 110 Job Corps Centers located throughout the U.S., tested the effects of vocational training, academic instruction, residential living, and other social and health education services on a large sample of eligible disadvantaged youths.	2	A	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Alternative Youth Employment Strategies Project	This demonstration, conducted from 1980 to 1982 in Miami, Albuquerque, and New York City, tested the effects of various combinations of work experience, training, and placement assistance on a large sample of low-income unemployed youth.	2	A	-	D
JOBSTART	This demonstration, conducted between 1985 and 1992 in 13 different cities tested the effects of education, training, and throughout the U.S., supportive services on a large sample of economically disadvantaged school dropouts.	2	A	-	-
Center for Employment Training (CET) Replication Study	This ongoing experiment, which began in 1995, tests whether the CET, which was originally tested in San Jose, can achieve success elsewhere. The CET model focuses on occupational training and offers training for a selection of occupations based on area employer demand. Basic skills needs of the clients are linked to the occupational training classes. The program is full time, with peer training and behavioral expectations that reflect the workplace. There is an emphasis on job placement, and clients (economically disadvantaged youth age 18 to 22) are considered to have graduated when they have found employment.	2	A	-	-
Young Families Can Project	This demonstration, conducted from 1987 through 1990 in Phoenix, tested the effects of case-management and counseling on a medium-sized sample of teenage mothers.	2	B	3	-
Minority Female Single-Parent Demonstration (MFSP)	This demonstration, conducted from 1984 to 1987 in Atlanta, San Jose, Providence, and the District of Columbia, tested the effects of education and training on a large sample of minority female single parents.	2	C	-	-
Youth Services and Conservation Corps	This demonstration, conducted 1993-94 in Washington State, New York City, Miami, and Santa Clara County (CA), tested the effects of a standard youth corps program—a nonresidential program employing groups of 5 to 10 young people in community service projects—on a medium-sized sample of out-of-school, 18- to 25-year-old youths.	2	D	-	-
Public Versus Private Sector Jobs Demonstration Project	This demonstration, conducted from 1978 to 1980 in five different states, tested the effects of subsidized work experience with either public or private employers on a large sample of low-income youths.	2	D	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Cambridge (MA) Job Search Voucher Program	This demonstration, conducted from 1980 to 1982, tested the effects of job search assistance and a wage supplement on a medium-sized sample of low-income youth.	2	I	-	-
Summer Experimental Youth Transition Project	This demonstration, conducted in 1980 in Baltimore, tested the effects of different forms of subsidized job search on a medium-sized sample of graduating low-income high school seniors.	2	I	-	-
Cambridge (MA) Job Factory	This demonstration, conducted from 1979 to 1980, tested the effects of job clubs, supervised job search assistance, and a bonus for finding a job quickly on a medium-sized sample of unemployed youth.	2	I	-	L
Wilkes-Barre (PA) Yes Workshop Program	This demonstration, conducted from 1979 to 1980, tested the effects of career counseling and job search assistance on a medium-sized sample of unemployed low-income youth.	2	I	-	-
Wage Subsidy Variation Experiment	This demonstration, conducted in 1980 in Detroit, tested the effects of a wage subsidy paid to employers on a medium-sized sample of private sector businesses. It measured the impact on employer agreement to participate in youth employment.	2	K	-	-
Wilkes-Barre (PA) Job Search Voucher Project	This demonstration, conducted in 1981, tested the effects of employer subsidies on a medium-sized sample of area employers. It measured the impact on employer employment of low-income youth.	2	K	-	-
Career Advancement Voucher Demonstration Project	This demonstration, conducted from 1979 to 1980 in five cities, all in different states, tested the effects of financial support for postsecondary education on a medium-sized sample of low-income youth.	2	M	-	-
Work First Profiling Demonstration	This demonstration, conducted from 1998 to 2000 in Michigan, tested the effects of a referral system in which clients were referred to various service providers based on a statistical employability assessment, on a large group of unemployed welfare recipients.	3	A	-	-
New York State Comprehensive Employment Opportunity Support Centers Program (CEOSC)	This demonstration, conducted from 1987 to 1993, tested the effects of training, job-search assistance, and supportive services on a medium-sized sample of single Aid to Families with Dependent Children (AFDC) parents of preschool children.	3	B	4	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Wisconsin Learnfare	This demonstration, conducted from 1990 to 1995, tested the effects of sanctions for inadequate enrollment or school attendance on a large sample of teenage AFDC recipients and their families.	3	B	-	-
Teenage Parent Home Visitor Services Demonstration	This demonstration, conducted from 1995 to 1997 in Chicago, Dayton, and Portland, tested the effects of mandated paraprofessional home visitor services on a large sample of teenage parents receiving AFDC.	3	B	-	-
Single-Parent Economic Independence Demonstration (SPEID)	This demonstration, conducted from 1988 to 1990 in Utah, tested the effects of unpaid internships and mentoring on a medium-sized sample of single parents. The random assignment design was not faithfully implemented, as program dropouts were removed from the treatment group in analysis.	3	B	-	D
AFDC Job Counselors	This demonstration, conducted from 1976 to 1979 in Michigan, tested the effects of employment counseling and placement services on a large sample of “inactive” Work Incentive Program (WIN) clients. Subjects were followed for 90 days.	3	B	-	I
West Virginia Community Work Experience Program (CWEP)	This quasi-experimental demonstration, conducted from 1983 through 1986, tested the effects of saturation Community Work Experience on a large sample of AFDC-U recipients (unemployed heads of two-parent families).	3	B	-	D
Baltimore Options Program (Maryland Employment Initiatives)	This demonstration, conducted from 1982 to 1985, tested the effects of education, training, job search, and work experience on a large sample of AFDC recipients.	3	B	-	-
San Diego Job Search and Work Experience Demonstration	This demonstration, conducted from 1982 to 1985, tested the effects of mandatory job placement assistance and work experience on a large sample of AFDC recipients.	3	B	-	-
Washington Community Work Experience Program	This work demonstration, conducted from 1982 to 1983, tested the effects of experience and job search assistance on a small sample of AFDC recipients.	3	B	-	-
Arkansas Work Program	This demonstration, conducted from 1983 to 1985, tested the effects of mandatory job search and work experience on a medium-sized sample of AFDC recipients.	3	B	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Maine Training Opportunities in the Private Sector (TOPS)	This demonstration, conducted from 1983 to 1987, tested the effects of vocational training and work experience on a medium-sized sample of unemployed single AFDC recipients.	3	B	-	-
Virginia Employment Services Program	This demonstration, conducted from 1983 to 1985, tested the effects of mandatory job search and work experience on a large sample of AFDC recipients.	3	B	-	-
Florida Trade Welfare for Work	This demonstration, conducted from 1984 to 1986, tested the effects of subsidized work experience and on-the-job training on a large sample of AFDC recipients.	3	B	-	-
New Jersey Grant Diversion Project	This demonstration, conducted from 1984 to 1987, tested the effects of subsidized on-the-job training on a large sample of AFDC recipients.	3	B	-	-
Cook County Job Search and Work Experience	This demonstration, conducted from 1985 to 1987, tested the effects of mandatory training and work experience on a large sample of AFDC recipients.	3	B	-	-
Saturation Work Initiative Model (SWIM)	This demonstration, conducted from 1985 to 1988 in San Diego, tested the effects of mandatory job search assistance and work experience on a large sample of AFDC recipients.	3	B	-	-
Pennsylvania Saturation Work Program (PSWP)	This demonstration, conducted from 1985 to 1987, tested the effects of intensive employment services on a large sample of welfare recipients.	3	B	-	-
Wisconsin Welfare Employment Experiment—Rock County	This demonstration, conducted from 1987 to 1990, tested the effects of Wisconsin’s new welfare program on a medium-sized sample of AFDC recipients. The Rock County experiment was part of a statewide demonstration, but was the only site to use random assignment in its design. Twenty-nine other counties tested the new program using a matched-comparison design.	3	B	-	A
Child Assistance Program (CAP)	This demonstration, conducted from 1988 to 1995 in three counties in New York State, tested the effects of financial incentives and intensive case management on a large sample of AFDC recipients.	3	B	-	-
Ohio Transitions to Independence Demonstration—JOBS	This demonstration, conducted from 1989 to 1992, tested the effects of employment and training participation on a large sample of welfare recipients. The JOBS Demonstration was one component of a larger evaluation, which also included the Work Choice Demonstration.	3	B	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Ohio Transitions to Independence Demonstration—Work Choice	This demonstration, conducted from 1989 to 1990, tested the effects of voluntary employment and training services, and extended transitional Medicaid and child care payments on a large sample of single-parent welfare recipients with children under age 6. This was one component of a larger evaluation which included the JOBS Demonstration.	3	B	-	-
Opportunity Knocks Program	This demonstration, conducted from 1989 through 1992 in DuPage County, Illinois, tested the effects of intensive case management on a small sample of low-income single-parent families.	3	B	-	-
AFDC Jobs Participants: Paths Toward Self-Sufficiency	This demonstration, conducted between 1990 and 1992 in Lincoln, Nebraska, tested the effects of in-home case management on a medium-sized sample of AFDC recipients.	3	B	-	-
Project Independence—Florida	This demonstration, conducted from 1990 to 1993, tested the effects of education, training, and enhanced support services on a large sample of AFDC recipients.	3	B	-	-
Partnership for Hope	This demonstration, conducted from 1990 to 1992 in Whatcom County, Washington, tested the effects of a case management approach, integrating family support services, mentoring, and job preparation training on a small sample of unemployed individuals qualifying for public assistance and considered at risk for child abuse.	3	B	-	-
National Evaluation of Welfare-to-Work Strategies	(Formerly known as Job Opportunities and Basic Skills Training [JOBS] Program Evaluation) This demonstration, conducted from 1991 to 1999 in seven sites in six different states, tested the effects of employment-focused strategies and education-focused strategies on a large sample of welfare recipients.	3	B	-	-
Link-Up	This demonstration, conducted between 1992 and 1994 in eight counties in California’s San Joaquin Valley, tested the effects of waiving the 100-hour rule for a large group of AFDC-U recipients.	3	B	-	-
To Strengthen Michigan Families	This demonstration, conducted from 1992 to 1996, tested the effects of a social contract requirement, financial incentives, and broadened AFDC eligibility on a large group of AFDC and State Family Assistance (SFA) recipients.	3	B	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Utah Single-Parent Employment Demo Program (SPED)	This demonstration, conducted from January 1993 to June 1996, tested the effects of active diversion, mandatory participation in work-related activities, repeal of the 100-hours rule, and other work incentives on a large sample of AFDC applicants.	3	B	-	-
California Work Pays Demonstration Program (CWDP)	This demonstration, conducted from January 1993 to June 1996, tested the effects of raising the earnings incentives and reducing the cash grant on a large sample of one- and two-parent AFDC families.	3	B	-	-
New Jersey Family Development Program	This demonstration, conducted from October 1993 to 1996, tested the effects of selected changes in welfare policy—a “family cap”, a higher earnings disregard, mandatory participation in education, training, and job search, an extension of Medicaid benefits, and strengthened sanctions—on a large sample of AFDC recipients.	3	B	-	-
Colorado Personal Responsibility and Employment Program	This demonstration, conducted between 1994 and 1997, tested the effects of consolidated services, mandatory employment services, and financial incentives on a large sample of AFDC recipients.	3	B	-	-
Minnesota Family Investment Program	This demonstration, conducted between 1994 and 1998, tested the effects of financial incentives and mandatory training on a large sample of welfare recipients and new applicants.	3	B	-	-
Vermont Welfare Restructuring Project	This demonstration, conducted from 1994 to 2001, tested the effects of financial incentives and work requirements on a large sample of welfare recipients.	3	B	-	-
Florida Family Transition Program	This demonstration, conducted from 1994 to 1999, tested the effects of time limits and earnings disregards on a large sample of families receiving welfare.	3	B	-	-
LA Jobs First GAIN (Greater Avenues for Independence) Evaluation	This demonstration, conducted from 1995 to 1996, tested Los Angeles (LA) County’s new work first program on a large sample of welfare recipients.	3	B	-	-
The Indiana Welfare Reform Evaluation	This demonstration, conducted from 1995 to 1999, tested the effects of welfare reform on two large samples of AFDC recipients.	3	B	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Virginia Independence Program	This demonstration, conducted from 1995 to 1997, tested the effects of eligibility reforms, time limits, work requirements, and work and savings incentives on a large sample of AFDC recipients.	3	B	-	-
Delaware: A Better Chance	This demonstration, conducted from 1995 to 1997, tested the effects of employment assistance, income incentives, and personal responsibility requirements on a large sample of welfare recipients in the State of Delaware.	3	B	-	-
Connecticut Jobs First	This demonstration, conducted from 1996 to 2002, tested the effect of a time limit, earnings disregard, and mandatory job search activities on a large sample of AFDC recipients.	3	B	-	-
Achieving Change for Texans (ACT)	This complex demonstration assessed the impact of welfare reform policies by comparing the outcomes of 44,852 TANF cases assigned to experimental or control groups in one of three experiments: (1) The Responsibilities, Employment and Resources (RER)–Non-Choices (This demonstration, conducted from January 1997 to September 2000, tested the effects of expanded eligibility and expanded caretaker responsibilities on a large sample of families on TANF), (2) Time Limits Pilot (This demonstration, conducted from June 1996 through September 2000, tested the effects of time-limiting benefits on a large sample of families receiving AFDC, and later on TANF), and (3) RER Choices (This demonstration, conducted from June 1996 through September 2000, tested the combined effects of time limits and RER provisions on a large sample of welfare recipients). The demonstration was designed to measure the impacts of these policies separately and in combination with each other.	3	B	-	-
Wisconsin Self-Sufficiency First/Pay for Performance Program	This demonstration, conducted from 1996 to 1997, tested the effects of mandatory job search or other employment-related activity on a large sample of AFDC families.	3	B	-	-
W-2 Child Support Demonstration Evaluation	This demonstration, conducted from 1998 to 2000 in Wisconsin, tested a liberalized pass-through of child support payments on a large sample of families receiving welfare.	3	B	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Food Stamp Work Registration and Job Search Demonstration	This demonstration, conducted from 1981 to 1983 in 21 sites located throughout the country, tested the effects of eight models of mandatory job search on a large sample of applicants for food stamps.	3	B	-	-
Food Stamp Employment and Training Program	This demonstration, conducted in 1988 in 53 sites in 23 states, tested the effects of the Food Stamp Employment and Training Program on a large sample of food stamp recipients.	3	B	-	-
New Visions College as a Job Advancement Strategy	This ongoing experiment, which began in 1997 is testing academic instruction and career guidance for TANF recipients who enter a 24-week core program of academic instruction and career guidance. The program offers classes in: remedial math, English, and reading; basic computer skills; and career life guidance which concentrates on critical thinking and problem-solving skills, as well as study and communication skills, needed at college and the workplace.	3	B	-	-
Welfare-to-Work Housing Voucher Demonstration	This ongoing demonstration, which began in 2000, is testing access to tenant-based rental assistance for families who have received Temporary Assistance for Needy Families (TANF) in the previous 2 years or who are TANF-eligible.	3	B	-	-
Ohio Learning, Earning, and Parenting Program (LEAP)	This demonstration, conducted from 1989 to 1995, tested the effects of applying bonuses and sanctions to the benefit checks of a large sample of custodial or pregnant teens on welfare in order to encourage school attendance.	3	B	-	-
Massachusetts Work Experience Program	This demonstration, conducted from 1978 to 1979, tested the effects of a mandatory work experience program, assisted job search, and a lower tax rate on earnings on a medium-sized sample of unemployed fathers receiving AFDC.	3	B	-	-
Louisville Immediate Job Search Assistance Experiment	This demonstration, conducted from 1978 to 1980, tested the efforts of immediate employment-assistance services and a counselor-directed job search program on a large sample of new female WIN registrants.	3	B	-	-
WIN Services to Volunteer	These two demonstrations, conducted from 1978 to 1979 in Denver, tested the efforts of a special recruitment effort and enriched employment assistance services in a large sample of female recipients of AFDC whose children were 5 years of age or younger.	3	B	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Louisville Group Job Search Experiment	This demonstration, conducted from 1980 to 1981, tested the effects of immediate employment-assistance services and incentive payments on a medium-sized sample of female WIN registrants.	3	B	-	-
Wisconsin Earned Income Disregard Demonstration	This demonstration, conducted from 1989 to 1992, tested the effects of a more generous income disregard on a large sample of AFDC recipients.	3	B	-	-
Alabama Avenues to Self-Sufficiency Through Employment and Training Services (ASSETS)	This quasi-experimental demonstration, conducted from July 1990 to June 1994, tested the effects of: (1) broadening the population subject to participation requirements in employment and training activities, and child support enforcement (CSE), efforts and (2) combining food stamps with AFDC in one cash grant for a large sample of AFDC and food stamp recipients.	3	B	-	-
Washington State Intensive Applicant Employment Services Evaluation	This quasi-experimental demonstration, conducted from 1982 to 1983, tested the effects of offering immediate services to AFDC applicants and extending work search requirements on a large sample of AFDC applicants.	3	B	-	-
West Virginia Community Work Experience Demonstration	This demonstration, conducted from 1983 to 1986, tested the effects of mandatory work experience on a large sample of AFDC recipients.	3	B	-	-
Evaluation of Iowa Family Investment Program	This demonstration, conducted between 1993 and 1999, tested the effects of eligibility reforms, mandatory employment services, and financial incentives on a large sample of AFDC recipients	3	B	-	-
Teenage Parent Demonstration	This demonstration, conducted between 1987 and 1991 in Chicago, Newark, and Camden, tested the effects of case management, and employment and training services on a large sample of teenage mothers receiving welfare.	3	B	-	-
CAL-Learn	This demonstration, conducted from 1994 to 1999 in California, tested the effects of financial incentives and intensive case management on a large sample of pregnant and parenting teenagers on welfare.	3	B	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Greater Avenues for Independence (GAIN)	This demonstration, conducted from 1988 to 1990 in six California counties, tested the effects of basic education, job search, and skills training on a large sample of AFDC recipients.	3	B	-	-
Washington State Family Independence Program (FIP)	This quasi-experimental demonstration, conducted between 1988 and 1993, tested the effects of financial incentives (bonuses) that were intended to encourage training and work on a large sample of AFDC recipients.	3	B	-	-
Child Day Care Recycling Fund Experiment—North Carolina	This demonstration, conducted from 1989 to 1990, tested the effects of immediate, guaranteed, subsidized day care on a medium-sized sample of AFDC recipients.	3	B	-	-
Community Group Participation and Housing Supplementation Demonstration (Bethel Self-Sufficiency Program)	This demonstration, conducted from 1989 to 1994 in a Chicago neighborhood, tested the effects of providing education, employment, and training services through a Community-Based Organization.	3	B	-	-
Cash Incentives in a Self-Sufficiency Program—Montgomery County, Maryland	This demonstration, conducted from 1989 to 1991, tested the effects of cash incentive payments on a medium-sized sample of AFDC recipients.	3	B	-	-
Illinois Youth Employment and Training Initiative (YETI)	This demonstration, conducted from 1994 to 1997, tested the effects of providing social and supplemental vocational skills on a medium-sized sample of students from TANF families.	3	B	-	-
New Jersey Substance Abuse Research Demonstration (SARD)	This ongoing experiment, which began in 1999, randomly assigns TANF recipients who are found to be in need of substance abuse treatment to one of two treatments: (1) Care Coordination (CC) which represents the standard substance abuse treatment available for welfare recipients in New Jersey, and 2) Intensive Case Management (ICM) which adds intensive case management and contingency interventions to the standard treatment.	3	B	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
National Supported Work Demonstration	This demonstration, conducted from 1975 to 1980 in 12 sites located throughout the U.S., tested the effects of supported work experience on a large sample of AFDC recipients, ex-offenders, substance abusers, and high school dropouts.	3	D	4,2	-
Homemaker-Home Health Aide Demonstrations	This demonstration, conducted from 1983 to 1986 in seven states, tested the effects of having AFDC clients work as aides in the homes of the elderly/impaired on large samples of both the elderly/impaired and recipients of AFDC.	3	D	-	-
WIN Job-Finding Clubs	This demonstration, conducted from 1976 to 1978 in New York City, New Brunswick (NJ), Milwaukee, Wichita, and Tacoma, tested the effects of mandatory supervised group job search training and assistance on a medium-sized sample of Work Incentive Program registrants.	3	I	-	-
Dayton Wage Subsidy Voucher Experiment	This demonstration, conducted from 1980 to 1981, tested the effects of a tax credit voucher for employers or a direct cash subsidy to employers on a medium-sized sample of welfare recipients.	3	K	-	-
Wisconsin Targeted Jobs Tax Credit (TJTC) Experiment	This demonstration, conducted in the early 1980s, tested the effects of a tax credit for employers on a medium-sized sample of WIN and TJTC eligible welfare recipients.	3	K	-	-
Training in Community Living	This demonstration, conducted from 1972 to 1976 in Madison, Wisconsin, tested the effects of on-call care and case management by trained staff members on a small sample of adults with mental disabilities.	4	C	-	-
Structured Training and Employment Transitional Services Demonstration (STETS)	This demonstration, conducted from 1981 to 1984 in Cincinnati, Los Angeles, New York City, St. Paul, and Tucson, tested the effects of training, work placement, and followup services on a medium-sized sample of young adults with low IQs.	4	C	-	-
Transitional Employment Training Demonstration (TETD)	This demonstration, conducted from 1985 to 1988 in 12 sites located throughout the country, tested the effects of job placement, on-the-job training, and followup services on a medium-sized sample of adults with mental retardation.	4	C	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Project Network	This demonstration, conducted from June 1992 to March 1995 in eight sites located throughout the U.S., tested the effectiveness of providing case and referral management services to a large sample of Social Security Disability Insurance (SSDI) recipients and disabled Supplemental Security Income (SSI) beneficiaries and applicants. In a followup study, earnings impacts were updated for 6 post-enrollment years, and SSDI benefit receipt information was updated for 7 post-enrollment years for SSDI beneficiaries based on administrative records.	4	C	-	-
Consumer-Operated Services Program	This ongoing experiment, which began in 2001 and is targeted at persons with mental illness, is testing consumer-operated services including vocational and housing programs, drop-in centers, peer counseling, case management services, crisis alternatives to hospitalization, advocacy training, and business ventures.	4	C	-	-
Employment Intervention Demonstration Program (EIDP)	This ongoing demonstration program, which began in 1995, is a set of eight related demonstrations, testing the effects of combined vocational rehabilitation with clinical services and supports on mental health consumers	4	C	-	-
State Partnership Initiatives (SPI)	SPI is a set of four related ongoing demonstrations which began in 1998. It is testing the effects of benefits counseling and other work support services on Supplemental Security Income (SSI) and Social Security Disability Income (SSDI) recipients.	4	C	-	-
Phoenix, Arizona Homeless Study	This ongoing demonstration, which began in 2001, is testing an intensive team case management approach using the principles of motivational learning for homeless women with children who have a mental health or substance abuse disorder.	4	C	-	-
Job Path	This demonstration, conducted from 1978 to 1980, tested the effects of subsidized and supervised work assignments on a small sample of adults with mental retardation.	4	D	-	-
Wild Cat Experiment	This demonstration, conducted from 1972 to 1978 in New York City, tested the effects of supported work experience on a medium-sized sample of adult substance abusers.	4	D	-	-
Living Insurance for Ex-Offenders (LIFE)	This demonstration, conducted from 1972 to 1975 in Baltimore, tested the effects of temporary financial assistance and job placement services on a medium-sized sample of male ex-offenders.	4	I	-	L

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Carbondale (IL) Handicapped Job-Finding Club	This demonstration, conducted from 1974 to 1975, tested the effects of supervised group job search assistance on a medium-sized sample of handicapped unemployed individuals.	4	I	-	-
Job Seekers' Workshop	This demonstration, conducted from 1976 to 1979 in San Francisco, tested the effects of an employment-related workshop and modified job club on small samples of heroin abusers.	4	I	-	-
Transition Project	This demonstration, conducted from 1980 to 1981 in New York City, tested the effects of job orientation meetings on a medium-sized sample of substance abusers.	4	I	-	-
Court Employment Project	This demonstration, conducted from 1977 to 1979 in New York City, tested the effects of case dismissal in exchange for acceptance of counseling and employment-related services on a medium-sized sample of individuals charged with felonies.	4	I	-	-
Homeless Employment Partnership (HEP)	This demonstration, conducted from 1989 through 1992 in Tacoma, Washington, tested the effects of intensive case management and employment services on a medium-sized sample of homeless men.	4	I	-	-
Alternative Schools	This demonstration, conducted from 1988 to 1990 in seven different sites, tested the effects of enrolling high-risk students in alternative high schools.	5	A	-	-
High-Risk Youth Demonstration Project—Yolo County, California	This demonstration, conducted between 1989 and 1991, tested the effects of a self-esteem-building/motivational curriculum on a small sample of high-risk youth.	5	A	-	-
Quantum Opportunities Program (QOP)—Pilot	This demonstration, conducted from 1989 to 1993 in San Antonio, Philadelphia, Saginaw (MI), and Oklahoma City, tested the effects of a comprehensive package of educational services, case management, and financial incentives on a medium-sized group of low-income high school students.	5	A	-	-
Quantum Opportunity Program (QOP) Demonstration	This demonstration, conducted from July 1995 through September 2001 in Cleveland, Fort Worth, Houston, Memphis, Washington, D.C., Philadelphia, and Yakima, tested the effects of intensive and comprehensive services on a large sample of youth with low grades entering high schools with high dropout rates.	5	A	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Austin Youth Leadership Development	This demonstration, conducted from October 1991 to September 1993, tested the effects of enhanced services on a sample of at-risk youth.	5	C	2	-
School Dropout Demonstration Assistance Program	This demonstration, conducted from September 1992 to June 1995 in cities located throughout the country, tested the effectiveness of varying dropout prevention programs on a large sample of at-risk youth	5	C	-	-
Career Academies	This ongoing demonstration, which began in 1993, is testing: (1) a school within-a-school organizational structure to create a supportive and personalized learning environment, (2) curriculum that combines academic and career or technical courses, and (3) partnerships with local employers to increase career awareness and provide work-based learning opportunities. The initial focus of keeping high school students at high risk of dropping out was later broadened to include preparing high-performing and high-risk students for college and employment.	5	C	-	-
Upward Bound	This is an ongoing evaluation of the Upward Bound Program, which began in 1994. Enrollees in Upward Bound consist of high school students from low-income families without a history of college attendance. They are offered traditional academic instruction, tutoring, mentoring, counseling, career planning, cultural programs, college planning activities, and intensive summer instruction.	5	C	-	-
Summer Training and Education Program (STEP)	This demonstration, conducted from 1985 to 1988 in Boston, Fresno, San Diego, Portland, and Seattle, tested the effects of a combined work and educational remediation program on a large sample of low-income youth.	5	I	-	-
Employment Retention and Advancement Project	This ongoing demonstration project, which began in 2000, consists of 15 separate experiments at 23 sites for current and former TANF recipients. The interventions use case management as a means of delivering other services which may include education or training, financial incentives to remain on the job, career planning, rehabilitation services, or job search assistance, depending on the demonstration.	6	C	-	-
Denver Post-Placement Services Project	This demonstration, conducted from 1980 to 1982, tested the effects of post-placement case management on a medium-sized sample of AFDC recipients who had recently found employment.	6	C	-	-

Table 2. List of Experiments, Descriptions, and Classifications (continued)

Experiment	Description	Target group	Program area	Other target groups	Other program area
Post-Employment Services Demonstration	This demonstration, conducted between the spring of 1994 and the fall of 1996 in Chicago, Portland, San Antonio, and Riverside County (CA), tested the effectiveness of providing case management services to a large sample of newly employed welfare recipients.	6	C	-	-
Emergency Food and Homelessness Intervention Project	This demonstration, conducted between 1988 and 1990 in Lincoln, Nebraska, tested the effects of in-home case management on a medium-sized sample of low-income families.	6	C	-	-
Parents' Fair Share Demonstration	This demonstration, conducted between 1994 and 1996 in seven cities located throughout the U.S., tested the effects of employment and training services and modified child support enforcement on a large sample of low-income non-custodial fathers who were delinquent in child support payments.	6	C	-	-
Operation INC (Incubator for New Companies)	This demonstration, conducted from 1989 to 1991 in Appleton, Missouri, tested the effects of a self-employment and small business startup program on a small sample of low-income individuals.	6	F	1	-
Wisconsin New Hope Project	This demonstration, conducted from 1994 to 1997, tested a program of earnings supplements and other work supports in Wisconsin on a large sample of low-income individuals.	6	L	1	-
Wisconsin Medical Assistance Extension Demonstration	This demonstration, conducted from 1989 to 1990, tested the effect of extending transitional medical assistance benefits on a large sample of ex-recipients of AFDC.	6	L	-	-
Illinois Career Advancement Project	This demonstration, part of the Self-Sufficiency Demonstration, was conducted between 1989 and 1993. It tested the effects of financial assistance for educational programs on a large sample of former AFDC recipients. It was not successfully implemented as planned, ending 1 year early due to minimal participation.	6	M	-	-

Appendix B: Summaries of Selected Social Experiments

Note: These summaries are extracted from the Third Edition of *The Digest of Social Experiments* (Greenberg and Shroder 2004).

CARBONDALE JOB-FINDING CLUB

SUMMARY: This demonstration, conducted in 1973, tested the effects of supervised group job search assistance on a small sample of the unemployed. Subjects were followed for three months.

TIME FRAME: 1973; data collected, 1973; final report, 1975.

TREATMENTS TESTED:

Controls. No treatment.

Experimentals. Daily group meetings to teach job search methods and develop positive job search attitudes through group reinforcement. Subjects received supervision in job search until successful. Elements of the treatment included the buddy system, secretarial services for resumes and letters of recommendation, a telephone bank, and job leads from other clients.

Experimentals were matched one-for-one with controls by an overall criterion of probable employability based on age, sex, race, education, marital status, desired position and salary level, number of dependents, and current financial resources. Once matched, a coin flip determined which member of the pair would be a control and which an experimental.

OUTCOME OF INTEREST: Employment.

SAMPLE SIZE: Experimentals, 60; controls, 60.

TARGET POPULATION: Unemployed persons not receiving unemployment benefits

MAJOR FINDINGS:

	Exper imentals	rols	Cont
Employed within three months of beginning treatment	92%		60%
Employed (more than 20 hours/week) within two months of beginning treatment	90%		55%
Mean starting wage	\$2.73	1	\$2.0
Median time until job found (includes those who did not find jobs during the period)	14 days	days	53

Note: These figures exclude experimentals who attended less than five sessions and their matched controls. No data are available on those excluded. All differences shown are statistically significant.

TIME TRENDS IN FINDINGS: Results are only reported through three months of treatment.

DESIGN ISSUES:

1. Results may be biased by self-selection, since only data on those who chose to attend five or more sessions were presented. The matching process did not control for self-selection except to the degree that the observed variables captured it.
2. The sample was small and quite heterogeneous.
3. Displacement bias seems very likely.
4. Success of the Job Club method probably varies with the size of the informal (unadvertised) job market, and this may vary among communities.

GENERALIZABILITY: Findings are striking, since Carbondale was a high-unemployment community. The sample excluded persons receiving unemployment insurance benefits because the project report authors believed some individuals were likely to lack motivation to find employment until their benefits ran out. They suggested, however, that if participation were a condition of receiving benefits, it would motivate job search in this group too.

FUNDING SOURCE: Illinois Department of Mental Health

TREATMENT ADMINISTRATOR: Anna Mental Health Center. Key personnel: Nathan H. Azrin.

POLICY EFFECTS: The Carbondale Job Finding Club demonstration is one of a number of job club experiments. Partially, based on the positive findings from these experiments, job clubs are now widely used.

DATA SOURCES: Impact evaluation was based on survey data. We have no information about a public-use file for this demonstration.

INFORMATION SOURCE: Nathan H. Azrin, T. Flores, and S. J. Kaplan, "Job-Finding Club: A Group-Assisted Program for Obtaining Employment," *Behavior Research and Therapy*, Vol. 13. 1975: 17–27.

CLAIMANT PLACEMENT AND WORK TEST DEMONSTRATION

SUMMARY: This demonstration, conducted in 1983, tested the effects of a deadline to register for work with the Employment Service, mandatory supervised job search, and job placement services on a large sample of Unemployment Insurance claimants. Subjects were followed for less than one year.

TIME FRAME: February–December 1983; data collected, same period; final report, July 31, 1984.

TREATMENTS TESTED:

Group A—Controls. The treatment of controls differed slightly from prior practice in that prior practice theoretically required Employment Service (ES) registration. Controls had no ES registration requirement and did not receive special job development efforts, although they could use the ES services voluntarily. They were required (as were experimentals) to come in periodically for eligibility reviews at the unemployment insurance (UI) office.

Group B. Improved work test, but regular Employment Service (ES) services. These experimentals were mailed notices coincidentally with their first week's UI check to come to the Employment Service office to register their availability for work. This practice differed from prior practice in that (1) the registration requirement was delayed so that those who never received a check did not come into ES offices, and (2) the registration was required as of some definite date. In general, failure to register would be taken as possible evidence of unavailability for work and, therefore, ineligibility for UI payments. New procedures were implemented to match ES and UI records so that this rule would be routinely enforced.

Group C. Improved work test and enhanced placement interviewer services. These experimentals received the same notice as group B. In addition, when the subject reported to the ES, an interviewer would attempt to develop a job for the subject unless he or she was a union member, was not job-ready, or was on layoff for some definite period. Group C subjects were also called in for a renewed job-placement attempt if still unemployed after nine weeks.

Group D. Improved work test, enhanced placement interviewer services, and job search workshops. In addition to group C experimentals, three-hour job search workshops were mandated for group D experimentals who were still receiving UI benefits four to five weeks after receiving enhanced placement interviewer services.

OUTCOMES OF INTEREST: (1) Employment and (2) UI payment reductions.

SAMPLE SIZE: Group A, 1,485; group B, 1,493; group C, 1,666; group D, 1,277.

TARGET POPULATION: New unemployment insurance (UI) claimants who had received an initial UI check, excluding those whose employers said they were on layoff for some definite period.

MAJOR FINDINGS:

	Group A	Group B	Group C	Group D
Percentage of subjects with nonmonetary determination (eligibility ruling)	5.8	13.4	16.7	18.4
Percentage of subjects with a denial (ruled ineligible)	4.2	7.4	9.2	8.7
1.	OLS confirms that all experimental treatments had a statistically significant positive effect on the denial rate compared with the experience of controls. This effect was substantially higher in <i>C</i> and <i>D</i> than in <i>B</i> , but differences between <i>C</i> and <i>D</i> were negligible.			
2.	Treatment effects on employment and wages as recorded in the UI wage reporting system were weak, inconsistent, and usually not statistically significant.			
3.	Effects on weeks of UI payments (coefficients of OLS dummy variables for treatment group; this measures the treatment effect on experimentals by comparison with controls):			

	Group B	Group C	Group D
Men	-0.83*	-1.15*	-1.14*
Women	-0.20	0.31	-0.15

*Statistical significance at 95 percent confidence level, on a two-tail test. OLS results control for cohort (week applying for UI).

Much of the difference between men and women results from the strong treatment effect on construction workers, who are mostly male. The experimental treatments reduced the weeks of UI received by male construction workers by about two weeks (the effects on female construction workers are about the same, but are not statistically significant because of small numbers). “A possible explanation for this result may relate to the casual, part-time nature of some construction employment (particularly during slack periods) and to the relatively low wage replacement rates that UI provides to construction workers.”

All experimental treatments were inexpensive: \$4.72, \$13.17, and \$17.58 per subject increments for treatments *B*, *C*, and *D*, respectively. The corresponding average reductions in UI payments were \$52.93, \$58.71, and \$73.14. All experimental treatments are therefore cost-effective, with the most cost-effective being treatment *D*.

TIME TRENDS IN FINDINGS: The variable representing cohort has a negative and statistically significant effect on weeks of unemployment. However, it is not possible to distinguish between the effect of falling unemployment rates and the effect of learning about new registration requirements for UI. Since experimental treatments differ in the times at which interventions occur, treatment effects are also indistinguishable from time effects.

DESIGN ISSUES:

1. Unemployment in Charleston fell from 8.9 percent to 6.6 percent in 1983. The improving economy reduced UI claims and increased job orders at the ES office.
2. If a claimant stated that he or she had failed to report as required because of illness or lack of transportation, there would have been a denial of UI on the grounds that the claimant was unable to work or was unavailable for work. On the other hand, claims that the summons to register at the ES had been lost in the mail were always accepted, even when there had been no difficulty receiving the UI check sent under a separate cover. The summons was sent under a separate cover only because of the experimental character of the demonstration (controls were not supposed to receive a call-in, and it

was apparently too complex to insert a call-in with some checks and not with others); experimental denial rates therefore slightly understate what could be expected upon implementation of the policy.

3. If the treatment effect is primarily centered on construction workers, it would not be surprising if no wage or employment effects were noted from UI data, because much of the industry is not covered or escapes reporting requirements.
4. In the cost-effectiveness analysis, instead of estimating the reduction in UI payments directly, the authors take the average UI weekly payment (\$96.24) and multiply it by the experimental treatment effect on weeks of receipt. This procedure fails to take into account the possibility that the part of the population whose behavior is changed by the treatment will not have the same average weekly benefit as the overall population.

GENERALIZABILITY: The project report authors were cautious about generalizability. They claimed that South Carolina Job Service/UI procedures are similar to those in most other states. The Charleston labor force has a higher percentage of blacks, however, than does the U.S. labor force. The importance of the construction industry, where much of the treatment impact was concentrated, varies across the country. State UI laws and regulations also vary; the maximum weekly payment in South Carolina, for example, was \$118.

FUNDING SOURCE: U.S. Department of Labor, Employment and Training Administration.

TREATMENT ADMINISTRATOR: South Carolina Employment Security Agency.

POLICY EFFECTS: No known effects.

DATA SOURCES: Impact evaluation based on administrative data, including data from the UI wage reporting system. We have no information about a public-use file for this demonstration.

INFORMATION SOURCES: Terry R. Johnson, Jennifer M. Pfister, Richard W. West, and Katherine P. Dickinson, "Design and Implementation of the Claimant Placement and Work Test Demonstration," SRI International, May 1984; Walter Corson, David A. Long, and Walter Nicholson, "Evaluation of the Charleston Claimant Placement and Work Test Demonstration," Mathematica Policy Research, July 31, 1984.

ILLINOIS UNEMPLOYMENT INSURANCE INCENTIVE EXPERIMENT

SUMMARY: This demonstration, conducted from 1984 to 1985, tested the effects of a reemployment bonus incentive on a large sample of unemployment insurance (UI) clients. Subjects were followed for up to one year.

TIME FRAME: Mid-1984 to early 1985; data collected, mid-1984 to mid-1985; final report, February 1987.

TREATMENTS TESTED:

1. Claimant experiment. A \$500 bonus was offered to an eligible claimant of unemployment insurance payments if he or she could find a job within 11 weeks and hold that job for four months.
2. Employer experiment. The same, except that the \$500 bonus would be paid to the claimant's employer.
3. Controls. Standard unemployment insurance eligibility.

OUTCOMES OF INTEREST: (1) Reductions in unemployment spells and (2) Net program savings.

SAMPLE SIZE: 4,186 claimant experimentals; 3,963 employer experimentals; 3,963 controls.

TARGET POPULATION: Persons who are (1) eligible for 26 weeks of unemployment insurance (UI) benefits; (2) between 20 and 55 years of age; and (3) registrants with Job Service offices who were not on definite layoff, ineligible for a union hiring hall, or not recent veterans or federal employees.

MAJOR FINDINGS:

1. A \$194 reduction in 52-week benefit payments to the average claimant experimental was found, by comparison with the average control. The comparable \$61 reduction for employer experimentals was not statistically significant.
2. A 1.15-week reduction in insured unemployment over 52 weeks was found for the average claimant experimental. The 0.36-week reduction for employer experimentals was not statistically significant.

3. The experimental treatments did not appear to curtail productive job search activity, because no statistically significant change in subsequent earnings was found between the control and experimental groups.
4. The employer experiment did result in a \$164 reduction per claimant in benefit payments to white women, which was statistically significant. The effects of the employer experimental treatment among blacks (male and female) and white males were not significantly different from zero. The claimant experimental response was statistically significant for whites of both sexes, but not for blacks of either sex.
5. Of claimant experimentals, 13.6 percent received a bonus; 25 percent qualified for one. Of employer experimentals, 2.8 percent obtained a bonus for their employers; 22.8 percent of them could have obtained one for their employers.
6. The ratio of benefit payments reductions to bonus cost for the claimant experiment was 2.32, which is statistically significant. The ratio for the employer experiment was 4.29 (not significant). If 100 percent of those eligible for bonuses had claimed them, the benefit–cost ratio would have been 1.26 for claimant experimentals and 0.53 for employer experimentals.

TIME TRENDS IN FINDINGS: Reductions in the initial unemployment spell may be slightly offset by increases in subsequent unemployment, but this is not statistically significant.

DESIGN ISSUES:

1. A potential displacement effect among nonparticipants in the program clearly exists.
2. Long-term market bias effects are conceivable if workers adjusted to a permanent bonus program by undergoing frequent short unemployment spells with accompanying state bonuses instead of infrequent long spells.
3. The low rate at which the bonuses were claimed by those who qualified for them is puzzling.

GENERALIZABILITY: The 22 sites involved in the experiment represented very diverse labor markets, so locality effects should not be present. Spiegelman noted that roughly the first half of workers subject to the experiment would have been eligible for extended (38-week) federal unemployment payments if their unemployed status had lasted longer than 26 weeks, whereas subsequent claimants were ineligible for them. Experimental response seems to have been stronger in the first group, but this was not conclusively established.

FUNDING SOURCES: Illinois Department of Employment Security (DES, using Wagner-Peiser grants); and W. E. Upjohn Institute for Employment Research.

TREATMENT ADMINISTRATOR: Illinois DES.

POLICY EFFECTS: This is one of four well-known reemployment bonus experiments. Because findings from these experiments tended to suggest that bonuses did not result in savings for the UI system, they tended to discourage adoption of bonus plans. However, data from the UI reemployment bonus experiments aided the development of profiling schemes used nationally in administering the UI program, as well as other programs.

DATA SOURCES: Impact evaluation based on administrative data, including data from the UI wage reporting system. We have no information about a public-use file for this demonstration.

INFORMATION SOURCES: Stephen A. Woodbury and Robert G. Spiegelman, "Bonuses to Workers and Employers to Reduce Unemployment: Randomized Trials in Illinois," *American Economic Review*, September 1987: 513–30. Woodbury and Spiegelman, "The Illinois Unemployment Insurance Incentive Experiments," final report to Illinois Department of Employment Security, W. E. Upjohn Institute, February 1987.

WASHINGTON AND MASSACHUSETTS UNEMPLOYMENT INSURANCE SELF-EMPLOYMENT WORK SEARCH DEMONSTRATIONS

SUMMARY: This demonstration, conducted from 1989 to 1991, tested the effects of business start-up services, including financial assistance, counseling, and workshop sessions, on a large sample of the recently unemployed. Subjects were followed for three years.

TIME FRAME: Demonstration period—Washington State: September 1989–March 1991; Massachusetts: May 1990–April 1993; data collected, same period; final report, December 1994.

TREATMENTS TESTED: Washington State: Subjects were offered four business training sessions; they continued to receive regular unemployment insurance (UI) payments and, in addition, were eligible for a lump-sum payment after achieving five program milestones (e.g., completion of training sessions and development of an acceptable business plan). Massachusetts: Subjects attended a one-day training session, an individual counseling session, and six workshop sessions on a variety of business topics. They continued to receive regular UI benefits and were exempted from UI work search requirements. Controls in both states remained eligible for regular UI benefits.

OUTCOMES OF INTEREST: (1) Self-employment, (2) Combined self- and wage and salary employment, and (3) Earnings

SAMPLE SIZE: Washington: treatment, 755; control, 752. Massachusetts: treatment, 614; control, 608.

TARGET POPULATION: New UI claimants were targeted. Filers of interstate claims, filers who were employer-attached (i.e., on standby to return to their former employer), and those who were under 18 were excluded. Massachusetts also excluded new claimants with a low predicted probability of exhausting UI benefits.

MAJOR FINDINGS: Results based on the longer of two observation periods (33 months in Washington, 31 months in Massachusetts).

1. Treatment group members in both states were much more likely than controls to have had a self-employment experience (+22 percent in Washington; +12 percent in Massachusetts) and to have spent more time per year in self-employment (+2 months in Washington; +0.8 months in Massachusetts). The Washington group was more likely to be still self-employed at the time of the second follow-up survey (the experimental–control difference was not statistically significant).
2. Annual self-employment earnings increased significantly in Washington (+\$1,675). These earnings also increased in Massachusetts, though not significantly.
3. In Washington, claimants' likelihood of working in wage and salary employment and their earnings from that employment were significantly reduced. Wage and salary earnings for the Massachusetts group increased significantly.
4. Combining self- and wage employment, the treatment groups in both states significantly increased their likelihood of employment and time employed when compared to control groups. Further, a significant increase in total annual earnings was seen in Massachusetts (+\$3,053).
5. The Washington treatment had a significant positive impact on the employment of nonparticipants (family and non-family employees).
6. The Washington demonstration significantly increased total program benefit payments. Taking into account both regular UI benefits and the lump-sum payment, total benefits were increased by approximately \$1,000. In contrast, the Massachusetts demonstration significantly reduced receipt of total benefits by nearly \$900.
7. The Massachusetts treatment generated large net gains from each of several benefit–cost perspectives (participant, nonparticipant, society, and government). The Washington treatment generated net gains from the perspective of participants and society, but resulted in net losses from nonparticipants and government perspectives.

TIME TRENDS IN FINDINGS: In Massachusetts, treatment group members were more likely than controls to have had a self-employment experience, but there was no difference in the percentage of self-employment after 33 months (at the time of the second follow-up survey). No time trends were reported for the Washington group.

DESIGN ISSUES:

1. In both states, only a small fraction (3.6 percent in Washington, 1.9 percent in Massachusetts) of targeted UI claimants met the initial requirements of attending an orientation meeting and submitting an application. Random assignment occurred at this point.
2. Massachusetts had legislative requirements to focus on UI claimants who were likely to exhaust their UI benefits. Sample selection was based on a statistical model that predicted claimants' likelihood of

benefit exhaustion. Those with low predicted probability were eliminated from the target group. There were no comparable requirements in Washington.

GENERALIZABILITY: Interest in self-employment is concentrated in a specific subgroup of the unemployed (males, ages 36–55, better educated, and with experience in professional, managerial, and technical occupations). Washington used a purposive site selection method based on an index of representativeness. Selected sites had the characteristics and diversity to enhance generalizability. Massachusetts' sites were self-selected, though they represented a wide geographic distribution and mix of salient characteristics.

FUNDING SOURCE: U.S. Department of Labor, Employment and Training Administration, Unemployment Insurance Service.

TREATMENT ADMINISTRATORS: Massachusetts—Department of Employment and Training (DET); Bonnie Dallinger, project manager. Washington—State Employment Security Department (ESD); Judy Johnson, project manager.

POLICY EFFECTS: According to the researchers, this demonstration had a “great impact” on the policy debate. It was cited by Congress, as part of the North American Free Trade Agreement (NAFTA), in the decision to authorize states to implement self-employment programs.

DATA SOURCES: Public use data are available through the U.S. Department of Labor.

INFORMATION SOURCE: Jacob M. Benus, Terry R. Johnson, Michelle Wood, Neelima Grover, and Theodore Shen, Self-Employment Programs—A New Reemployment Strategy: Final Impact Analysis of the Washington and Massachusetts Self-Employment Demonstrations, U.S. Department of Labor, Employment and Training Administration, Unemployment Insurance Occasional Paper 95-4, Washington, 1995.

JOBSTART

SUMMARY: This demonstration, conducted between 1985 and 1992, tested the effects of education, training, and supportive services on a large sample of economically disadvantaged school dropouts. Subjects were followed for four years.

TIME FRAME: Demonstration period, 1985–1988; data collected through 1992; final report, October 1993.

TREATMENTS TESTED: Education and vocational training, support services, and job placement assistance. Support services included assistance with transportation, child care, counseling, and incentive payments.

OUTCOMES OF INTEREST: (1) Educational attainment, (2) Employment, (3) Earnings, and (4) Welfare receipt.

SAMPLE SIZE: Total, 2,312 (1,941 used for analysis—see Design Issue #3). Treatment, 1,163; control, 1,149.

TARGET POPULATION: Economically disadvantaged school dropouts ages 17–21 who read below the eighth grade level and were eligible for Job Training Partnership Act (JTPA) Title IIA programs or the Job Corps. Some sites screened out youth with problems that the program was not equipped to handle. These problems included: emotional problems, drug and alcohol abuse, health problems, unstable living conditions, poor motivation, and those who were likely to prove dangerous or disruptive.

MAJOR FINDINGS:

1. Significantly more treatment group youths completed high school or equivalent as compared to the control group (42 percent versus 28.6 percent). This impact was fairly large for all subgroups studied.
2. Employment was significantly greater for the control group in the first year of follow-up. In the second year, significantly more treatment group youths were employed. There were no significant differences in the third and fourth years regarding this variable.
3. Similarly, control group members earned more in the first two years after follow-up whereas the treatment group earned more in the third and fourth years. This finding was statistically significant only for year one.
4. Few significant findings are found in subgroup analyses regarding employment. Trends are similar to those for the full sample.

5. For women who had no children when the program started, those in the treatment group were less likely to have children and to receive AFDC payments at follow-up than were women in the control group. For custodial mothers (those with children at the start of the program), this finding was reversed; those in the treatment group were more likely to have more children at follow-up. However, this finding was not statistically significant.
6. There is some indication that JOBSTART led to a reduction in criminal activity (arrests and drug use), although impacts are generally not significant.
7. The San Jose site had higher earnings impacts than any other site. This finding was significant at the .01 level. The reasons for this are unclear. At the San Jose site, training and placement efforts were closely linked to the labor market, education and training efforts were coordinated, and the program had a clear organizational mission. Any or all of these factors may have affected program impacts.

TIME TRENDS IN FINDINGS: Treatment group payoffs did not usually occur until after year two.

DESIGN ISSUES:

1. The intake process was quite lengthy at some sites. Some youth who were eventually assigned to the treatment group had already found other opportunities or had lost interest.
2. At some sites, assessment was done after random assignment. This led to the inclusion of some participants with reading levels higher than the eighth grade.
3. Roughly 84 percent of the sample responded to the 48-month follow-up survey. These 1,941 youths were used for the analysis. There were significant differences between responders to the follow-up survey and nonresponders, although responders were not more likely to be experimentals than controls. Since nonresponse was randomly distributed, it does not alter the expected values of the adjusted mean outcomes, and thus does not bias impacts.

GENERALIZABILITY: Designed to generalize to the entire nation. Site selection was not a probability sample. Sites were chosen for their ability to meet program guidelines, assemble operational Funding, and yield sufficient sample members. At some sites, the assessment process screened out youths with significant problems. Despite this, the JOBSTART sample still appears to be slightly more disadvantaged than the majority of youth served nationwide by JTPA Title IIA programs, but had slightly fewer barriers to employment than those in Job Corps.

FUNDING SOURCES: Funding was through a consortium of the U.S. Department of Labor, the Rockefeller Foundation, the Ford Foundation, Charles Stewart Mott Foundation, the William and Flora Hewlett Foundation, National Commission for Employment Policy, the AT&T Foundation, Exxon Corporation, ARCO Foundation, Aetna Foundation, the Chase Manhattan Bank, and Stuart Foundations.

TREATMENT ADMINISTRATOR: Local Service Delivery Areas (SDAs) at the 13 sites. The final report (p. 20) lists all treatment administrators.

POLICY EFFECTS: None known.

DATA SOURCES: Public use data are available from MDRC.

INFORMATION SOURCES: George Cave, Hans Bos, Fred Doolittle, and Cyril Toussaint, *JOBSTART: Final Report on a Program for School Dropouts*. MDRC, October 1993.

MINORITY FEMALE SINGLE-PARENT DEMONSTRATION (MFSP)

SUMMARY: This demonstration, conducted from 1984 to 1987, tested the effects of education and training on a large sample of minority female single parents. Subjects were followed for up to five years.

TIME FRAME: Demonstration period, November 1984–December 1987. Data collected over same period, with follow-up of up to five years (San Jose only).

TREATMENTS TESTED: The four centers used different approaches, but all included assessment, education, job training, and support services such as counseling and child-care assistance. The control group was not eligible to receive services at the centers, but could seek them elsewhere.

OUTCOMES OF INTEREST: (1) Earnings, (2) Employment, and (3) Welfare receipt.

SAMPLE SIZE: Total applicants, 3,965. Total used in analysis, 3,175; treatment, 1,841; control, 1,334. By site (treatment; control): Atlanta, Georgia, 373; 299; San Jose, California, 440; 329; Providence, Rhode Island, 346; 163; and Washington, D.C., 682; 543.

TARGET POPULATION: Minority, single mothers.

MAJOR FINDINGS: The San Jose CET, unlike the three other sites, immediately provided job-specific skill training to all trainees and integrated basic literacy and mathematics skills into the job training curriculum. Participants at this site demonstrated the greatest employment and earnings gains. In contrast, the other sites provided more traditional remedial education and training.

1. The CET program generated the greatest gains in employment and earnings, and these lasted over the full 60-month period (only CET had a 5-year follow-up; the other three sites were followed for only 30 months). During the fifth year after program application, treatment group members earned an average of \$95 per month more than did control group members, a statistically significant impact equal to 17 percent of control group mean earnings.
2. CET's impact on welfare receipt, at 30 and 60 months, was small and not statistically significant.
3. Over the 5-year period, CET produced a positive return from the perspectives both of society and of program participants (in 1986 dollars, \$975 and \$2,500 per participant, respectively). From a government-budget perspective, costs exceeded benefits by about \$1,600 per participant.
4. WOW generated modest and significant gains in employment rates, but not in average earnings. Reductions in welfare receipts were small and not significant.
5. The investment in WOW did not produce a positive return from either the social or the government-budget perspective.
6. The projects at AUL and OIC had no significant impacts on any of the outcome variables.
7. None of the programs had significant long-term impacts on measures of psychological well-being, including depression and locus of control, or on fertility or marriage behavior.

TIME TRENDS IN FINDINGS: At the 30-month follow-up in the CET program, all treatment group members showed significant gains in employment and earnings. At five years, only sample members with 12 or more years of schooling showed significant gains. Impacts on general equivalency diploma (GED) attainment were significant for the treatment group at 30 months, but this effect had dissipated at 60 months.

DESIGN ISSUES: The Minority Female Single-Parent (MFSP) demonstration consisted of four projects in different locations. The impacts of each project may be due to the characteristics of each site. Differences in impacts across sites should be treated cautiously.

GENERALIZABILITY: Designed to generalize to the entire nation. MFSP projects were operated by community-based organizations independent from the local welfare offices and reach a somewhat broader clientele than state-run JOBS programs. MFSP participation was entirely voluntary. In contrast, JOBS programs are closely linked to the local welfare office and serve only welfare recipients. Participation is mandatory for some AFDC recipients. MFSP participants were more disadvantaged than national samples of minority single mothers; on average, they were younger, had younger children, had less employment experience, and less education.

FUNDING SOURCE: The Rockefeller Foundation.

TREATMENT ADMINISTRATORS: Atlanta Urban League (AUL)—key personnel: Lyndon Wade and Edna Crenshaw; Center for Employment and Training (CET)—key personnel: Russ Hershey, Carmen Ponce, and Carmen Placida; Opportunities Industrialization Center of Rhode Island (OIC)—key personnel: Michael van Listen and Kathy May; and Wider Opportunities for Women (WOW)—key personnel: Cindy Marino and Barbara Maoris.

POLICY EFFECTS: The Rockefeller Foundation was active in disseminating the results of the CET findings and entering them into the policy debate. The U.S. Department of Labor has funded CET to replicate their model in other programs. There is also currently a Labor Department evaluation of CET replication projects. (See Appendix II, at the end of this volume, for a summary of this ongoing experiment.)

DATA SOURCE: Impact evaluation largely based on administrative data. Public use data are available through Mathematica Policy Research.

INFORMATION SOURCES: John Burkhart, Ana Rangarajan, Anne Gordon, and Ellen Kisker, *Evaluation of the Minority Female Single Parent Demonstration: Summary Report*, Mathematica Policy Research, October 1992; Amy Zambrowski and Anne Gordon, *Evaluation of the Minority Female Single Parent Demonstration: Fifth-Year Impacts at CET*, Mathematica Policy Research, December 1993.

NATIONAL JOB CORPS STUDY

SUMMARY: This study, conducted from 1994 to 1996, tested the effects of vocational training, academic instruction, residential living and other social and health education services on a large sample of eligible disadvantaged youths. Subjects were followed for four years.

TIME FRAME: Data collected from November 1994 to February 1996; and at 12, 30 and 48 months after random assignment.

TREATMENTS TESTED: All eligible Job Corps applicants nationwide were randomly assigned to either the program group (permitted to enroll in Job Corps services) or the control group (not permitted to enroll in Job Corps services for three years). Job Corps centers provided vocational training, academic instruction, health care, social skills training, and counseling. The number and types of trades for which vocational training were provided varied from site to site. Placement agencies provided post-program services for all program participants for a period of six-months. The majority of participants (88%) lived at the centers, which are usually located in rural areas at some distance from the participants' urban homes. Nonresidential participants were primarily females with children.

OUTCOME OF INTEREST: Employment, earnings, education and job training, welfare receipt, criminal behavior, drug use, health factors and household status.

SAMPLE SIZE: Total, 15,386. Control, 5,977. Program group, 9,409.

TARGET POPULATION: Disadvantaged men and women between the ages of 16 and 24.

MAJOR FINDINGS:

Impact on education and training enrollment		
	Percentage enrolled	Average percentage of total weeks spent in education or training
Program group	93	24
Control group	72	18
Estimated impact per eligible applicant	21	6

Impact on credential receipt		
	Percentage who received GED	Percentage who received vocational certificate
Program group	42	38
Control group	27	15
Estimated impact	15	23

1. Job Corps had no effect on college attendance, illegal drug use, child support payments, custody of children, or fertility rates.
2. Placement services were insufficient. Few program participants received assistance in securing employment in the six-month post-program period.
3. Approximately two years following random assignment, the program group surpassed the control group in earnings. In the fourth year there was a 12% earnings gain for program participants. However, there was no increase in employment and earnings for Hispanic program participants.
4. The program group received an estimated \$639 less per participant in combined cash assistance and food stamps benefits than did the control group.
5. The Program group had 16% reduction in the overall number of arrests. The impact was greater on reducing arrests for less serious crime.
6. The benefit-cost analysis suggests that the program returns about \$2 to society for every dollar spent, due to the combination of increased participant output in years 2–4; reduced use of other programs and services; reduced crime and the output produced during program vocational training.

TIME TRENDS IN FINDINGS: Early in the follow-up period, the control group had higher earnings and rates of employment because many of the program group members were enrolled in Job Corps services. Beginning about two years after random assignment, the average earnings for program group members gradually grew to be higher than the average earnings of control group members and persisted through the fourth year follow up period.

DESIGN ISSUES: None apparent.

GENERALIZABILITY: This study is based on a fully national random sample of all eligible applicants to Job Corps with services delivered nationwide. Since many members of the control group received training and education from outside sources, the findings of the study should be interpreted as the incremental effects of Job Corps Services.

FUNDING SOURCE: U.S. Department of Labor (DOL), Employment and Training Administration.

TREATMENT ADMINISTRATOR: Local Job Corps centers.

POLICY EFFECTS: Findings from this experiment are fairly recent and do not appear to have strongly influenced policy, except to have helped maintain political support for the program.

DATA SOURCE: Impact analysis mainly relied on survey data. Public use data available through the Department of Labor.

INFORMATION SOURCE: John Burghardt, Peter Z. Schochet, Sheena McConnell, Terry Johnson, R. Mark Gritz, Steven Glazerman, John Homrighausen, and Russell Jackson, *Does Job Corps Work? Summary of the National Job Corps Study*, 2001, and the reports cited therein. This report was produced by Mathematica Policy Research, Inc. under contract with the U.S. Department of Labor, Employment and Training Administration.

OHIO'S LEARNING, EARNING AND PARENTING PROGRAM (LEAP)

SUMMARY: This demonstration, conducted from 1989 to 1995, tested the effects of applying bonuses and sanctions to the benefit checks of a large sample of custodial or pregnant teens on welfare in order to encourage school attendance. Subjects were followed for four years.

TIME FRAME: Demonstration period, August 1989-September 1991; data collection occurred during a four year follow-up from August 1990-September 1995; final report in July 1997.

TREATMENTS TESTED:

1. The experimentals faced a financial incentive tied to school attendance and sanctions for non-compliance. The benefit increase was \$62 for school enrollment with an additional \$62 for regular monthly attendance. The sanction was an equal amount and was assessed for unacceptable absences or non-enrollment. Temporary exemptions were allowed for medical, childcare, and transportation reasons. Experimentals also received case management and child care and transportation services. Cleveland had an enhanced version of LEAP, with school-based services such as childcare, intensive case management, special GED classes, and other services.
2. The control group was not eligible for any component of LEAP except its childcare services on an as available basis. They had access to state run programs that offered parenting classes, but little else.

OUTCOMES OF INTEREST: (1) School enrollment, (2) Attendance, (3) Graduation, (4) GED receipt, (5) Welfare receipt, (6) Employment, and (7) Earnings.

SAMPLE SIZE: Four-year follow-up: 4151 teens. 3479 program and 672 controls. Some analyses use a smaller three year sub-sample.

TARGET POPULATION: Pregnant or custodial teen parents under 20 years of age who were receiving welfare.

MAJOR FINDINGS:

1. Program succeeded in reaching 93% of eligible teens.
2. LEAP significantly increased school enrollment and attendance. Initial school status was an important predictor of program success; those who were enrolled when the program started had better outcomes.
3. Ohio Department of Human Services and taxpayers' benefits outweighed costs. LEAP teens experienced a net loss of \$1,110 over the four years. This was calculated as the difference in gains in earnings and tax credits, and losses due to reductions in AFDC, Food Stamps and Medicaid eligibility.
4. Enhanced program in Cleveland had some impact on high school graduation rates for those teens initially enrolled.

Asterisks indicate significant findings. ***=1 percent **=5 percent *=10 percent

Measure	Treatment	Control	Impact
Full sample			
In the 12 months after random assignment			
Number of months enrolled in high school	4.8	4.2	0.6**
Number of months enrolled in a GED program	1.3	0.8	0.5***
In the 3 years after random assignment			
Ever completed grade 11 (%)	50.0	45.4	4.6*

Ever completed high school (%)	22.9	23.5	-0.6
Ever received a GED (%)	11.1	8.4	2.7
Ever employed			
Year 2 (%)	43.8	40.6	3.1*
Year 3 (%)	51.3	49.8	1.4
Year 4 (%)	61.0	59.6	1.4
Total earnings, years 1-4 (\$)	4,405	4,293	112
Amount of AFDC received, years 3 and 4 (\$)	5,185	5,459	-275**

Sample members enrolled in high school or in a GED program at random assignment

In the 12 months after random assignment			
Number of months enrolled in high school	7.3	6.6	0.7*
Number of months enrolled in a GED program	0.9	0.7	0.3
In the 3 years after random assignment			
Ever completed grade 11 (%)	60.6	58.1	2.5
Ever completed high school (%)	35.6	34.2	1.4
Ever received a GED (%)	10.0	4.4	5.6**
Ever employed			
Year 2 (%)	46.4	39.7	6.7***
Year 3 (%)	55.7	54.7	1.0
Year 4 (%)	65.1	60.5	4.6*
Total earnings, years 1-4 (\$)	4,862	4,319	544
Amount of AFDC received, years 3 and 4 (\$)	5,181	5,497	-316**

Sample members not enrolled in high school or in a GED program at random assignment

In the 12 months after random assignment			
Number of months enrolled in high school	1.5	1.0	0.5*
Number of months enrolled in a GED program	1.7	0.9	0.8***
In the 3 years after random assignment			
Ever completed grade 11 (%)	35.8	28.0	7.8*
Ever completed high school (%)	6.7	7.8	-1.1
Ever received a GED (%)	12.0	14.3	-2.3
Ever employed			
Year 2 (%)	41.0	42.2	-1.2
Year 3 (%)	46.3	44.0	2.4
Year 4 (%)	56.3	58.7	-2.6
Total earnings, years 1-4 (\$)	3,930	4,271	-341
Amount of AFDC received, years 3 & 4 (\$)	5,172	5,395	-223

TIME TRENDS IN FINDINGS: No consistent time trend.

DESIGN ISSUES:

1. The research sample entered the program before LEAP was fully implemented. This may have resulted in conservative results being reported.
2. Teens identified problems at school as reason for non-attendance that LEAP program was not designed to address.

GENERALIZABILITY: Designed to be generalizable to the rest of the country. However the problem of identifying eligible teens, which created implementation difficulties in this program, could seriously hamper less technologically advanced systems.

FUNDING SOURCE: Ohio Department of Human Services. Additional funding: Ford Foundation, the Cleveland Foundation, BP America, the Treu-Mart Fund, the George Gund Foundation, the Procter & Gamble Fund and the U.S. Department of Health and Human Services.

TREATMENT ADMINISTRATOR: Ohio Department of Human Services. County offices.

POLICY EFFECTS: Policy effects not known.

DATA SOURCE: Impact analysis mainly based on administrative records. Public use data are available, see MDRC website at www.mdrc.org/tsch_publicuse.htm

INFORMATION SOURCES: Johannes Bos, Veronica Fellerath, *LEAP: Final Report on Ohio's Welfare Initiative to Improve School Attendance Among Teenage Parents*, Manpower Demonstration Research Corporation, August 1997.

David Long, Robert Wood, Hilary Kopp, *LEAP: The Educational Effects of LEAP and Enhanced Services in Cleveland*, Manpower Demonstration Research Corporation, October, 1994.

Dan Bloom, Veronica Fellerath, David Long, Robert Wood, *LEAP: Interim Findings on a Welfare Initiative to Improve School Attendance Among Teenage Parents*, Manpower Demonstration Research Corporation, May 1993.

SAN DIEGO JOB SEARCH AND WORK EXPERIENCE DEMONSTRATION

SUMMARY: This demonstration, conducted from 1982 to 1985, tested the effects of mandatory job placement assistance and work experience on a large sample of AFDC recipients. Subjects were followed for two years.

TIME FRAME: Demonstration period, October 1982–August 1983; data collected through March 1985; final report, February 1986.

TREATMENTS TESTED: Random assignment occurred at the time of application for AFDC. Impacts per applicant therefore include those who were found ineligible for AFDC (and were therefore ineligible for programs).

1. Experimental group 1: Job placement assistance on day of AFDC application, followed by three weeks of Job Club.
2. Experimental group 2: Same treatment as group 1, but if still unemployed at the end of 3 weeks, required to hold an unpaid work experience job at a public or nonprofit agency for up to 13 weeks. The hours of unpaid work were set by the family's welfare grant divided by the minimum wage.
3. Controls: Minimal WIN services.

OUTCOMES OF INTEREST: (1) Employment, (2) Earnings, and (3) AFDC reciprocity and payments.

SAMPLE SIZE: AFDC-R: group 1, 856; group 2, 1,502; controls, 873. AFDC-U: group 1, 831; group 2, 1,376; group 3, 813.

TARGET POPULATION: AFDC and AFDC-U applicants who were WIN-mandatory. Excluded were refugees, persons with language barriers, and applicants with children under age six.

MAJOR FINDINGS:

1. Work experience supervisors found the productivity of their assigned subjects roughly comparable to those of regular entry-level employees.
2. Sanctions were applied to from 4 percent to 8 percent of experimentals, but to 1 percent or less of controls.

Findings for AFDC-Regular:

1. Both experimental groups had significantly higher rates of employment (61 percent to 55 percent) over six quarters than the controls, but the group 1 (no Workfare) differential faded to insignificance after the third quarter.

2. The Workfare group had \$700 more in earnings per subject over six quarters than the control; the non-Workfare group had \$251 more in earnings, but this was not significant.
3. Neither treatment had much effect on AFDC reciprocity.
4. Over six quarters, the AFDC payments per Workfare experimental were \$288 less than those per control; payments to Job-Club-only experimentals were \$203 less than those per control, but this difference was not significant.
5. Program gains were largest among those in the sample who had no previous employment experience.

Findings for AFDC-U:

1. “For both program models, there were statistically significant and substantial reductions in welfare payments, but no significant impacts on the employment and earnings of AFDC-U applicants.” The Workfare subjects received on average \$530 less in AFDC payments over six quarters than did controls; the job-search-only subjects received \$470 less.
2. “Sanctioning rates were higher for experimentals than for controls, and those sanctioned faced larger grant reductions than did AFDC’s (regulars).”
3. “In general, mandating (Workfare) for AFDC-U’s did not improve program outcomes compared to those found for the Job Search program.” The benefit–cost analysis findings are that there were consistent, large net gains to taxpayers and government budgets for both programs and for both applicant groups. AFDC-R applicants also benefited financially from the Workfare treatment; the job-search-only treatment did not always have positive benefits. AFDC-U applicants were made worse off by the treatments, because they reduced their benefits without increasing their earnings.

TIME TRENDS IN FINDINGS: AFDC-R welfare savings from the Workfare treatment declined over time. AFDC-U welfare savings from the two treatments seem to decline gradually over time.

DESIGN ISSUES:

1. Site self-selection. Workfare is not politically acceptable everywhere.
2. The San Diego job market tightened over the course of the experiment. This affected the applicant mix. The study attempts to compare earlier with later cohorts to examine this.
3. The work experience programs seem to have been unusually well run, without any sense of make-work or strong client resentment detected in the process analysis. It is not clear that other counties could replicate that success.

GENERALIZABILITY: Generalization to the full AFDC population tested in the Saturation Work Initiative Model (SWIM) experiment.

FUNDING SOURCES: US Department of Health and Human Services; Ford Foundation, and State of California.

TREATMENT ADMINISTRATOR: State Employment Service Agency and the County Welfare Department. Key personnel: Ray Koenig and Joan Zinser.

POLICY EFFECTS: Reports on the State Welfare to Work Initiatives were frequently cited in Congress during the framing of the Family Support Act (FSA) of 1988. This demonstration also led to the SWIM experiment.

DATA SOURCE: Impact findings based on administrative data. Public-use access does not exist.

INFORMATION SOURCES: Barbara S. Goldman, Daniel Friedlander, and David A. Long, *California: Final Report on the San Diego Job Search and Work Experience Demonstration*, Manpower Demonstration Research Corporation, February 1986.

GREATER AVENUES FOR INDEPENDENCE (GAIN)

SUMMARY: This demonstration, conducted from 1988 to 1990, tested the effects of basic education, job search, and skills training on a large sample of AFDC recipients. Subjects were followed for up to four and one-half years.

TIME FRAME: Demonstration period, March 1988–June 1990; final Report, September 1994.

TREATMENTS TESTED: Experimentals were subject to the Greater Avenues for Independence (GAIN) participation mandate, which combined basic education, job search activities, assessments, skills training, and work experience. The control group was precluded from receiving those services from the program, but could seek other services in the community on their own.

OUTCOMES OF INTEREST: (1) Participation in employment-related activities, (2) Earnings, (3) Welfare receipts, and (4) Employment.

SAMPLE SIZE: Total, 32,751; treatment, 24,528; control, 8,223.

MAJOR FINDINGS: Effects varied by county.

1. Over the full four and one-half years, GAIN significantly increased the earnings for AFDC-FG (single parents) as compared to those of the control group. There was a total difference of \$2,527 per participant (24 percent). This is an all-county average, with each county weighted equally.
2. AFDC total benefit payments were, on average, \$1,365 (7 percent) lower for the AFDC-FG treatment group than for the controls.
3. Earnings gains and AFDC-FG benefit reductions were greatest for Riverside County (+44 percent and -1.5 percent, respectively). Earnings gains and AFDC-U reductions for Riverside County were +15 percent and -14.3 percent, respectively. Riverside's staff placed much more emphasis on moving registrants into the labor market quickly than did staff in any other county. Riverside staff attempted to communicate a strong message to all registrants that employment was the central goal of the program. In addition, the county's management established job placement standards as one means of assessing staff performance.
4. Employment and welfare savings impacts for the AFDC-U (two-parent families) group were less striking than for the AFDC-FG group, though they were statistically significant. Earnings for treatment group members averaged 12 percent higher than those of their control group counterparts.

TIME TRENDS IN FINDINGS: The earnings impacts grew progressively stronger over time, whereas the impacts on AFDC receipt tended to level off over time.

DESIGN ISSUES:

1. In order for group assignment to occur, registrants were required to appear for orientation and appraisal. One-third of mandatory registrants did not appear and thus were not part of the research sample. By the six-month follow-up, two-thirds of these no-shows had either left welfare or were officially excused.
2. Four of the six counties were able to enroll the full mandatory caseload. Funding levels did not permit this in Alameda and Los Angeles, where they chose to focus exclusively on long-term recipients. Different intake policies and differences in the general makeup of each county's local population yielded research samples that varied markedly in demographic composition.

GENERALIZABILITY: Designed to generalize to the entire state of California. The program was mandatory under pre-JOBS rules. Thus, single parents with children under the age of six were exempted. The six counties represent diverse geographical regions of the state, vary widely in local economic conditions and population characteristics, and constitute a mix of urban and rural areas. GAIN, as California's version of the JOBS program, existed statewide. Although participants in the six experimental counties were not strictly representative of the statewide population, over one-half of the entire state AFDC caseload lived in those counties.

FUNDING SOURCE: California State Department of Social Services (SDSS).

TREATMENT ADMINISTRATOR: SDSS. Key personnel: Eloise Anderson, Michael Genest, and Bruce Wagstaff, as well as the county welfare directors in the six experimental counties.

POLICY EFFECTS: The work-first approach used in the GAIN Riverside site and the positive findings for this site became widely known in the U.S. and elsewhere and has probably encouraged wide-spread adoption of work-first approaches.

DATA SOURCE: Impact analysis mainly based on administrative data. Public use data available through MDRC.

INFORMATION SOURCES: James Riccio, Daniel Friedlander, and Stephen Freedman, *GAIN: Benefits, Costs and Three Year Impacts of a Welfare to Work Program*, Manpower Demonstration Research Corporation, 1994.

**NATIONAL EVALUATION OF WELFARE-TO-WORK STRATEGIES
(Formerly known as Job Opportunities and Basic Skills Trainings [JOBS] Program Evaluation)**

SUMMARY: This demonstration, conducted from 1991 to 1999, tested the effects of employment-focused strategies and education-focused strategies on a large sample of welfare recipients. Subjects were followed for five years.

TIME FRAME: Enrollment period, June 199 –December 1994; data collected through December 1999.

TREATMENTS TESTED: Eleven programs, broadly defined as either employment-focused or education-focused, were tested in seven sites across the United States. Participation in all the programs was mandatory, although enforcement of this requirement varied among the tested programs.

		Employment-focused		Education-focused	
		Job search first	Varied first activity	Education or training first	
	Number of groups (including control)	High enforcement	High enforcement	High enforcement	Low enforcement
Atlanta	3	X		X	
Grand Rapids	3	X		X	
Riverside	3	X		X	
Portland	2		X		
Columbus	3			X Integrated X Traditional	
Detroit	2				X
Oklahoma City	2				X

In Atlanta, Grand Rapids, and Riverside, two programs operated simultaneously: 1) Labor Force Attachment (LFA) - an employment-focused program in which members were assigned to job clubs and encouraged to find work quickly; 2) Human Capital Development (HCD) – a mandatory, education-focused program in which most members were assigned to the class deemed most appropriate: GED preparation, adult basic skills, or English as a Second Language (ESL). In Columbus, two case management strategies were tested: (1) In the Traditional program, each recipient worked with two different staff members—one for income maintenance issues and one for employment and training issues. (2) In the Integrated program, recipients worked with only one staff member, who handled both aspects of their cases. In both programs, participation in Job Opportunity and Basic Skills (JOBS) was mandatory. In the Portland (employment-focused) program, some enrollees were first assigned to a short-term education program to attain GED certification. Program group members were encouraged to seek well-paying jobs. In all of the programs, control group members received income maintenance, but could not receive program services.

OUTCOMES OF INTEREST: Employment, earnings, welfare receipt, cost-effectiveness, and child well-being.

SAMPLE SIZE: Full sample, 44,569.

MAJOR FINDINGS:

1. Most control group members (ranging from 79 percent in Oklahoma City to 88 percent in Grand Rapids) worked at some point in the five -year follow-up period. Thus, there was little program effect on the percentage of treatment group members who “ever” worked.
2. Nearly all of the 11 programs increased employment and earnings. However, treatment group members without a high school diploma had larger overall impacts from the employment-focused programs than from the education-focused programs.
3. None of the programs made welfare recipients substantially better off financially. Individuals replaced welfare and Food Stamp dollars with dollars from earnings and Earned Income Tax Credits, but the total income of treatment group families did not increase above the low levels of the control group.
4. Portland’s program had the largest and most consistent effect. Relative to the control group, treatment group members had a 21 percent increase in employment duration and, on average, earned \$5000 more over the five-year period.
5. There were no significant program effects on employment in the two “low enforcement” sites (Detroit and Oklahoma City).
6. All programs reduced welfare expenditure relative to control levels, but the HCD approach did not produce added economic benefits relative to the LFA approach, and the HCD approach was 40% to 90% more expensive.
7. In Columbus, the impact on earnings was larger for the Integrated group members than for the Traditional group members. However, the impact was not statistically significant among those who had a high school diploma.
8. There were few program effects on measures of child-well being. The effects, when found, did not differ by HCD versus LFA strategy.

TIME TRENDS IN FINDINGS: Employment-focused programs produced effects immediately; education-focused programs did not produce effects until the second year following random assignment. Effects for most programs diminished during the last two years and were statistically insignificant by the end of year 5.

DESIGN ISSUES: In a few programs, a small portion of the control group members received services under the tested interventions during years 4 and 5.

GENERALIZABILITY: Given the diversity of sites, large samples, and consistency with results of other studies, the cross-site findings on earnings, welfare receipt, and income appear generally applicable to the population of welfare families.

FUNDING SOURCE: U.S. Department of Health and Human Services (HHS) with support from U.S. Department of Education.

TREATMENT ADMINISTRATOR: Local welfare agencies.

POLICY EFFECTS: The exceptionally positive findings have been widely discussed in policy circles.

DATA SOURCES: Impact findings are mainly based on administrative data. Data are available from the National Center for Health Statistics (NCHS). <http://aspe.hhs.gov/hps/NEWS>.

INFORMATION SOURCE: Gayle Hamilton, Stephen Freedman, Lisa Gennetian, Charles Michalopoulos, Johanna Walter, Diana Adams-Ciardullo, Anna Gassman-Pines, *National Evaluation of Welfare-to-Work Strategies: How Effective Are Different Welfare-to-Work Approaches? Five-Year Adult and Child Impacts for Eleven Programs*, Manpower Demonstration Research Corporation, November 2001. Stephen Freedman, *National Evaluation of Welfare-to-Work Strategies: Four-Year Impacts of Ten Programs on Employment Stability and Earnings Growth*, Manpower Demonstration Research Corporation, December 2000.

TEENAGE PARENT DEMONSTRATION

SUMMARY: This demonstration, conducted between 1987 and 1991, tested the effects of case management and employment and training services on a large sample of teenage mothers receiving welfare. Subjects were followed up to seven years.

TIME FRAME: Demonstration period, 1987 to 1991; data collected over same period and for 30-month and for 6-½ year follow up.

TREATMENTS TESTED: Individual case management. Participants were required to attend an initial series of personal skills workshops, followed by full-time (30 hours per week) participation in education, training and/or employment services. Sanctions for nonparticipation consisted of a reduction in the monthly AFDC grant. Childcare and transportation assistance were available.

OUTCOME OF INTEREST: 1) Employment, 2) earnings, 3) welfare receipt, 4) school attendance, 5) subsequent childbearing and, 6) child outcomes.

SAMPLE SIZE: Total, 5297; treatment, 2647; control, 2650.

TARGET POPULATION: Teenage mothers with only one child, receiving AFDC for the first time.

MAJOR FINDINGS:

Early results:

1. Participation in employment and, especially, education activities was significantly higher for the treatment group (79 percent) than for the control group (66 percent). By the end of the two-year follow-up period, a significantly higher percentage of treatment group members, compared with control group members, had spent some time in school. These differences were largest in Camden (46 percent versus 26 percent). In Chicago and Camden, significantly more treatment group members obtained a diploma or equivalent.
2. The demonstration produced consistent but modest impacts on employment outcomes. For both Chicago and Camden, there were positive significant treatment–control differences in the percentage ever employed (Camden, 6.5 percent, Chicago, 7 percent) and in the percentage of total demonstration period employed (Camden, 2.4 percent, Chicago, 3.5 percent). There were no significant differences for the Newark site on employment outcomes.
3. Average monthly earnings were significantly higher for the treatment group at the Chicago site only (\$24.40 higher than controls). In the New Jersey sites, the treatment groups had average monthly earnings of roughly \$21 over the control groups, but this was not statistically significant.
4. Welfare receipt was reduced by 7–8 percent in all three sites. Particularly in New Jersey, these reductions resulted from a combination of higher earnings and higher sanctions for noncompliance. Only the Chicago program produced a sustained significant reduction in the probability of receiving AFDC.

Follow-up results:

1. Early program impacts generally disappeared once the demonstration ended and the program services and sanctions for noncompliance were discontinued. (See Time Trends, below.)
2. All programs significantly increased school participation. Programs in Camden and Chicago also increased participation in vocational training significantly in the first phase.
3. In Newark, experimentals were significantly less likely to be living with a husband or partner (12 percent, compared with 18 percent of the controls).
4. There were little or no significant differences between treatment and control groups in children's cognitive and social-emotional well-being or physical health.

TIME TRENDS IN FINDINGS: The first phase of the demonstration showed that the programs increased rates school enrollment, job training, and employment over the 30-month evaluation period. Modest employment increases were accompanied by earning gains that reduced rates of welfare dependence. However, impacts on employment and welfare receipt did not last once the demonstration ended. Roughly 70 percent of the mothers in both the treatment and control groups were still receiving welfare at the time of the 6 1/2 year follow up. Overall, there were no long-term impacts on subsequent births, employment, earnings, educational attainment, or child development.

DESIGN ISSUES:

1. The evaluator reported that this project was very well designed and was implemented with few difficulties. The problems that existed related to how quickly the three sites implemented the project and to how faithfully they monitored it. One site was fairly slow and another site was inconsistent in monitoring the project.
2. Sample sizes for outcomes based on resurvey and retest data are substantially smaller than those based on records data. This was purposeful and not due to attrition and nonresponse (response to the surveys was roughly 85 percent). Evaluators attempted to survey and retest only 75 percent of the Chicago sample.

GENERALIZABILITY: Findings appear reasonably generalizable to urban teenage parents.

FUNDING SOURCE: U.S. Department of Health and Human Services (DHHS), Administration for Children and Families.

TREATMENT ADMINISTRATOR: Chicago, Project Advance; key personnel: Melba McCarty. Newark, Teen Progress; key personnel: Yvonne Johnson. Camden, Teen Progress; key personnel: Frank Ambrose.

POLICY EFFECTS: No known policy effects.

DATA SOURCE: Available through U.S. Department of Health and Human Services. Contact Reuben Snipper, Office of the Assistant Secretary for Planning and Evaluation.

INFORMATION SOURCES:

Rebecca Maynard, Walter Nicholson, and Anu Rangarajan, *Breaking the Cycle of Poverty: The Effectiveness of Mandatory Services for Welfare-Dependent Teenage Parents*. Mathematica Policy Research, December, 1993.

Ellen Eliason Kisker, Anu Rangarajan, and Kimberly Boller, *Moving Into Adulthood: Were the Impacts of Mandatory Programs for Welfare-Dependent Teenage Parents Sustained After the Programs Ended?* Mathematica Policy Research, February 1998.

POST EMPLOYMENT SERVICES DEMONSTRATION

SUMMARY: This demonstration, conducted between the spring of 1994 and the fall of 1996, tested the effectiveness of providing case-management services to a large sample of newly-employed welfare recipients. Subjects were followed for two years.

TREATMENTS TESTED: Post Employment Services Demonstration (PESD) case managers provided clients with (1) counseling and support, (2) job search assistance, (3) help with applying for benefits and resolving eligibility problems, (4) service referrals, and (5) payments for temporary work-related expenses. Control group members received the regular services provided for employed welfare recipients.

OUTCOMES OF INTEREST: (1) Employment, (2) Earnings, and (3) Welfare receipts.

SAMPLE SIZE: Chicago, Illinois—Experimentals, 552; controls, 993. Portland, Oregon—Experimentals, 425; controls, 379. Riverside, California—Experimentals, 500; controls, 1006. San Antonio, Texas—Experimentals, 386; controls, 392.

MAJOR FINDINGS:

1. The PESD treatment had little effect on employment, earnings, or welfare receipt.
2. Some clients needed minimal assistance or short-term assistance (e.g. finding child care) to remain employed. Other clients had serious or multiple barriers and required more assistance on an ongoing basis.
3. Although the PESD programs offered to provide clients with employer mediation to help them retain their jobs, and many clients experienced job-related problems that made it difficult for them to remain employed, most clients did not want their case managers to intervene on their behalf. The clients did not want their employers to know of their welfare connection and to feel that they required mediation assistance from others.
4. The PESD treatment was not more effective for some sub-groups than for others.

TIME TRENDS IN FINDINGS: None.

DESIGN ISSUES:

1. The impacts of the programs might have been reduced by the strong economic conditions that prevailed during the demonstration, the pioneering nature of the program, and the evolutionary character of the experiment.
2. The four sites had various levels of pre-existing services for welfare recipients who had recently been hired and different post-employment services for experimentals.
3. Case managers had little guidance on what their roles were, what services they were to deliver, and how they were to deliver services.
4. The PESD programs provided all clients with case management services, regardless of client needs.
5. The services received by control group members in Portland and Riverside were very similar to the services received by treatment group members at those sites.
6. The programs in Riverside and Portland adhered closely to the requirements of serving the target population. However, 25 to 35 percent of the clients at the San Antonio and Chicago sites were JOBS-exempt welfare recipients.

GENERALIZABILITY: Generalizable. The final report suggested that similar case-management-based programs working with like populations would not be more successful than the PESD.

FUNDING SOURCE: Funding was provided by grants from the Administration for Children and Families of the U.S. Department of Health and Human Services.

TREATMENT ADMINISTRATOR: The Illinois Department of Human Services was the lead state agency. David Gruenfelder was the project officer.

POLICY EFFECTS: Because case-management alone did not appear effective, a new experiment, the Employment Retention and Advancement Project, is now testing a range of different treatments to increase retention and advancement.

DATA SOURCE: Public use files not available.

INFORMATION SOURCES: J. Haimson, A. Hershey, and A. Rangarajan, *Providing Services to Promote Job Retention*, Mathematica Policy Research, October 1995; Anu Rangarajan and Tim Novak, *The Struggle to Sustain Employment: The Effectiveness of the Postemployment Services Demonstration, Final Report*, Mathematica Policy Research, April 22, 1999.

NATIONAL SUPPORTED WORK DEMONSTRATION

SUMMARY: This demonstration, conducted from 1975 to 1980, tested the effects of supported work experience on a large sample of AFDC recipients, ex-offenders, substance abusers, and high school dropouts. Subjects were followed for two years.

TIME FRAME: Demonstration March 1975–December 1978; data collected, April 1975–mid-1980; final report, 1981.

TREATMENTS TESTED: (1) Controls received no treatment. (2) Experimentals were offered employment in a structured work experience program involving peer group support, a graduated increase in work standards, and close sympathetic supervision, for 12 to 18 months. Local agencies contracted with Manpower Demonstration Research Corporation to employ the experimentals in a broad range of activities, with pay starting at the minimum wage (or slightly higher, depending on local market conditions), and bonuses and merit increases for workers who met increasing work standards. Agencies maintained a high ratio of supervisors to participants (1:8 to 1:12), and implemented different on-site methods for crew interaction and shared responsibility. Typical work activities were construction, building maintenance, and child day care.

OUTCOMES OF INTEREST: (1) Increases in post-treatment earnings; (2) Reductions in criminal activity; (3) Reductions in transfer payments; (4) Reductions in drug abuse.

SAMPLE SIZE: Experimentals, 3,214; controls, 3,402.

TARGET POPULATION: (1) Long-term recipients of AFDC (30 of last 36 months, no children under age six); (2) Ex-addicts following drug rehabilitation treatment (within past 6 months); (3) Ex-offenders, aged 18 or over, incarcerated within past 6 months; (4) Young school dropouts, aged 17–20, not in school past 6 months, at least 50 percent having delinquent or criminal records.

MAJOR FINDINGS:

1. Major positive effect on earnings of AFDC-recipient group.
2. Minor increase in earnings of ex-addicts, and major reduction in criminal activity.
3. No discernible effects on young dropouts.
4. No clear effects on ex-offenders.
5. Benefits exceeded costs for AFDC recipient and ex-addict groups by \$8,000 and \$4,000, respectively. Costs exceeded benefits for young dropouts. For ex-offenders, the bulk of findings show costs substantially exceeding benefits (experimentals were arrested more frequently), but a small-sample, three-month follow-up shows the reverse tendency.

TIME TRENDS IN FINDINGS: Earnings differences showed little decay over time among recipients of Aid to Families with Dependent Children (AFDC); criminal activity differences fell over time among ex-addicts.

DESIGN ISSUES:

1. Local organizations competed to win these contracts. Some projects were not funded, and one was discontinued for poor performance. Thus, self-selection of sites might bias the results.
2. Displacement effects could occur in two ways. First, the greatest impact seemed to be for AFDC recipients in periods of high unemployment. Second, the agencies competed for local government contracts, set up small businesses, and so on; and they might have displaced other businesses.
3. Both controls and experimentals underreported arrests. A research finding that underreporting seemed to be of the same magnitude between controls and experimentals is critical to results.

GENERALIZABILITY: The large number of sites and subjects adds power to the findings. However, site and contractor self-selection are certainly present.

FUNDING SOURCES: U.S. Department of Labor (DOL), Education and Training Administration; U.S. Department of Justice, Law Enforcement Assistance Administration; U.S. Department of Health and Human Services (DHHS), Office of Planning and Evaluation, National Institute on Drug Abuse (NIDA); U.S. Department of Housing and Urban Development, Office of Policy Development and Research; U.S. Department of Commerce, Economic Development Administration; and the Ford Foundation.

TREATMENT ADMINISTRATOR: Manpower Demonstration Research Corporation. Key personnel: Judith M. Gueron.

POLICY EFFECTS: Because of the high cost of the tested treatment and the fact that it only proved clearly successful for one of the four target groups, it has not come into wide-spread use.

DATA SOURCE: Impact findings are mainly based on survey data. Public-use access file exists; contact Manpower Demonstration Research Corporation.

INFORMATION SOURCES: Board of Directors, Manpower Demonstration Research Corporation (MDRC), *Summary and Findings of the National Supported Work Demonstration*, MDRC, 1980; Peter Kemper, David A. Long, and Craig Thornton, *The Supported Work Evaluation: Final Benefit–Cost Analysis*, MDRC, 1981.

NATIONAL JOB TRAINING PARTNERSHIP ACT (JTPA) STUDY

SUMMARY: This demonstration, conducted from 1987 to 1991, tested the effects of the Job Training Partnership Act (JTPA) Title II program's employment and training services on a large sample of economically disadvantaged adults and youths. Subjects were followed for 30 months.

TIME FRAME: Enrollment period, November 1987–September 1989; data collected through December 1991; final report, 1994.

TREATMENTS TESTED: Access to Title II-A services under the JTPA. Participants were divided into three groups by local staff according to which services were deemed appropriate. They were then randomly assigned to a treatment or control group for each service strategy. Specific services varied widely across sites, but could include the following:

	Service strategy		
	Classroom training group	On-the-job training (OJT)/job search assistance group	Other activities group
Specific program service			
Classroom training occupational skills	Yes	No	Yes
OJT	No	Yes	Yes
Job Search Assistance	Yes	Yes	Yes
Basic education	Yes	Yes	Yes
Work experience	Yes	Yes	Yes
Miscellaneous	Yes	Yes	Yes

The control group was not allowed to receive services for 18 months.

OUTCOMES OF INTEREST: (1) Earnings; (2) Employment; (3) Welfare receipt; and (4) Attainment of educational credentials and occupational competencies.

SAMPLE SIZE: Full experimental sample, 20,602. Classroom training group, 7,090; on-the-job training group, 7,412; and other activities group, 6,100. Sample size includes both treatment and control group members. On average, 68 percent of sample members were randomly assigned to treatment groups.

TARGET POPULATION: Eligible JTPA Title II adults and out-of-school youth. The study focused on four subgroups: adult women; adult men; female out-of-school youths; and male out-of-school youths.

MAJOR FINDINGS: Because the control group was able to receive employment and training services from non-JTPA providers, impacts reflect the incremental effect of JTPA services beyond what sample members could have accomplished without access to JTPA. Impacts were estimated separately by subgroups: adult men; adult women; female youth; male youth, non-arrestees; and male youth, arrestees.

Adults:

1. The treatment group received significantly ($p = .01$) more employment and training services than did the control group; on average, men received 169 more hours of service and women received 136 more hours.
2. For adult women, average earnings over the 30-month period following random assignment were \$1,176 (9.6 percent) greater for the treatment group than the control group. This is significant at the .01 level. For men, earnings were \$978 (5.3 percent) greater for the treatment group. This is significant at the .10 level.
3. Earnings gains came more from an increase in hours worked (an employment effect), than from an increase in average hourly earnings (a wage effect). This was especially true for women.
4. JTPA resulted in a substantial and statistically significant impact on the attainment of a high school credential (diploma or equivalent) for adult female school dropouts. The findings for adult males were also positive, although not statistically significant.
5. The greatest earnings impact was estimated for women in the OJT/JST and other activities subgroups.
6. For adult women, there was no significant program impact on Aid to Families with Dependent Children (AFDC) or food stamp receipt. For men, there was a small, but significant, increase in AFDC receipt for the treatment group.

Youth:

1. JTPA resulted in a significant increase in the amount of employment and training services for all categories of youth. (Female youth in the treatment group received, on average, 182 more hours than their control group counterparts; male youth non-arrestees received 175 more hours; and male youth arrestees received 127 more hours.)
2. There were no significant treatment—control group differences for the quarterly earnings of female youths and male youth non-arrestees. For male youth arrestees, there was a great discrepancy between survey data and data using unemployment wage records. The former suggests significantly less earnings for the treatment group. The wage record data suggest no significant treatment—control group difference.
3. JTPA had a significant positive effect on the attainment of a high school credential for female youths (7.7 percent more treatment group females, compared to their control group counterparts, had a high school diploma or GED 30 months after random assignment), but not for male youths.
4. No significant treatment—control group differences were found for welfare receipts for male or female youths.

TIME TRENDS IN FINDINGS: There was a gradual increase in the earnings of all adult participants—treatment and control group—over time.

DESIGN ISSUES:

1. Impacts are reported per assignee, but 34 percent of women and 38 percent of men in the treatment group did not participate in JTPA. Therefore, this reflects the impact of offering JTPA services, rather than receiving them. Impacts per participant were not estimated directly from the experimental data, but, rather, were inferred using an extension of the data. The estimates reported above are in terms of impact per assignee, rather than per participant.
2. Site selection was done on a voluntary basis. Service delivery areas (SDAs) were not mandated to participate. Many were reluctant because they feared political fallout from random assignment, they found the design too complex, or they could not obtain agreement among all local participants.
3. Because JTPA program staff often recommends more than one program service for an applicant, the study was designed to measure impacts of clusters of program services, not single services in isolation, such as classroom training, on-the-job training, or job search assistance.

4. Formal agreements with some of the SDAs excluded certain small groups of applicants from the study (and from random assignment) owing to logistical reasons, recruitment difficulties, and/or the nonvoluntary nature of certain applications.
5. A 2 to 1 ratio of treatment to control group members was used to minimize the number of persons that had to be turned away by local program staff.

GENERALIZABILITY: Site selection was not a probability sample, and the SDAs that volunteered to be part of the study may differ from the national population in unobservable ways. They differed in two observable ways: (1) No large, central cities were included owing to the decentralized nature of service in these locations; and (2) The study sites tended to emphasize classroom training and job search assistance more, and OJT and miscellaneous services less, than their counterparts nationally. However, the observable characteristics of program participants in the study sites were similar to the characteristics of participants nationwide.

FUNDING SOURCE: U.S. Department of Labor.

TREATMENT ADMINISTRATOR: Local service delivery areas (SDAs) in the 16 sites. See Orr et al. (1996: ix-x) for key personnel at each site.

POLICY EFFECTS: Following the results and recommendations of the National JTPA study, the U.S. Department of Labor proposed a 47 percent reduction in funding for Title II-C (the program for out-of-school youth) and a modest increase for Title II-A (the adult program). Congress responded by applying a large reduction for Title II-C and allowing the 11 percent increase in funding for the adult program to remain intact. Thus, findings appear to have had an impact on national policy.

DATA SOURCE: Both survey and administrative data used for impact analysis. Public use data are available through U.S. Department of Labor.

INFORMATION SOURCES: Larry L. Orr, Howard S. Bloom, Stephen H. Bell, Fred Doolittle, Winston Lin, and George Cave, *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study*, Urban Institute Press, 1996; Larry L. Orr, Howard S. Bloom, Stephen H. Bell, Winston Lin, George Cave, and Fred Doolittle, *The National JTPA Study: Impacts, Benefits, and Costs of Title II-A*, Abt Associates, 1994; Fred Doolittle and Linda Traeger, *Implementing the National JTPA Study*, Manpower Demonstration Research Corporation, 1990.

PROJECT NETWORK

SUMMARY: This demonstration, conducted from June 1992 to March 1995, tested the effectiveness of providing case- and referral-management services to a large sample of Social Security Disability Insurance (SSDI) recipients and disabled Supplemental Security Income (SSI) beneficiaries and applicants. Subjects were followed for two or three years. In a follow-up study earnings impacts were updated for 6 post-enrollment years, and SSDI benefit receipt information was updated for 7 post-enrollment years for SSDI beneficiaries based on administrative records.

TIME FRAME:

DEMO: Model 1 – June 1992 - June 1994; Model 2 – January 1993 - January 1995; Model 3 – February 1993 - February 1995 (New Hampshire), March 1993 - March 1995 (Richmond, Virginia); Model 4 – January 1993 - January 1995.

DATA COLLECTION: January 1990-December 1997. Administrative records update in 2002.

FINAL REPORT: March 22, 1999.

TREATMENTS TESTED: Treatment group members received case- or referral-management services sponsored by the Social Security Administration (SSA). These services included assessment of the individual, development of an employment plan for the individual, and identification of and arrangement for services that would be necessary to achieve the individual's employment plan. These services were provided to treatment group members by one of four different models, each of which operated for 24 months. Control group members were eligible for all non-NetWork services that were already available to them. Some of these services were quite similar to Project NetWork

services. Members of the treatment and control groups were also offered waivers of the SSDI and SSI rules that were viewed as disincentives to work.

In Model 1, case management services were provided by SSA staff.

In Model 2, case management services were provided by private rehabilitation organizations, which operated under contract to SSA.

In Model 3, case management services were provided in SSA offices by State Vocational Rehabilitation Agencies, which operated under contract to SSA.

Model 4 featured the provision of referral management services by SSA staff. The referral managers found services for clients that were already being provided in the community, including case management services. Each model was tested in 2 sites, and there were 8 sites in all.

OUTCOMES OF INTEREST: (1) Employment, (2) Earnings, (3) SSI and SSDI receipts, and (4) Health and well-being.

SAMPLE SIZE: Total, 8,248; treatment, 4,160; control, 4,088. Total Model 1, 1,899; treatment, 956; control, 943. Total Model 2, 2,112; treatment, 1,088; control, 1,024. Total Model 3, 2,214; treatment, 1,087; control, 1,127. Total Model 4, 2,023; treatment, 1,029; control, 994.

TARGET POPULATION: Beneficiaries of SSDI and disabled applicants for and recipients of SSI. Participants were all volunteers and were eligible for Project NetWork regardless of their age or the nature of their disabilities.

MAJOR FINDINGS:

1. All four models were able to enroll large numbers of participants and to provide them with rehabilitation and employment services “on a substantial scale.”
2. Treatment group members experienced a statistically significant 6-percentage point increase in the receipt of rehabilitation, employment, and training services.
3. Treatment group members experienced a statistically significant \$220 (11 percent) increase in their annual earnings for the first two years of the experiment. For SSDI beneficiaries earnings were subsequently followed up for six post-enrollment years, and persisted throughout this longer follow-up period, with an annual average of \$239 for the 6-year period (2002 dollars)³. Treatment group members did not experience a statistically significant reduction in SSI or SSDI receipt during the follow-up period for the initial evaluation. For SSDI beneficiaries, 84-month follow-up data did not indicate a statistically significant reduction in SSDI receipt either...
4. The Project NetWork evaluation did not detect statistically significant improvements in the health or well-being of treatment group members relative to control group members.
5. Project NetWork increased the average number of months employed by a statistically significant 3 weeks (21 percent) during the first two follow-up years.
6. A nonexperimental comparison indicated that the referral model had an estimated impact on earnings that was statistically significantly smaller than impacts in the other three models.
7. When asked if they would participate in Project NetWork if they had to make the choice over again, 77 percent of treatment group members said they would.
8. Project NetWork generated a net cost to taxpayers of \$2,019 per experimental.

TIME TRENDS IN FINDINGS: See items (2) and (3) under Major Findings.

DESIGN ISSUES:

1. There were “substantial delays” between the assessment of individuals and the development of the employment plans for the individuals. This delay caused some members of the treatment group to lose interest in the demonstration.

2. Managers in the case management models stated that they wished they had more training in the development of Individual Employment Plans for clients. Managers in the referral management model stated that they would have liked more training in the development of Individual Referral Plans.
3. Both treatment and control group members were offered various waivers to facilitate implementation, and therefore the estimated net impacts are conditional on the presence of the waivers. A nonexperimental evaluation of “waiver effects” was inconclusive. However, the process analysis showed that the waivers were not implemented as planned, suggesting that “waiver effects” were probably negligible during the evaluation period.

GENERALIZABILITY: Generalizable. However, prior to the demonstration, the employment rates of persons with disabilities were slightly higher than the national average in all sites. This may reflect a greater willingness of the local employers to hire persons with disabilities, which would bias the results upward.

FUNDING SOURCE: SSA

TREATMENT ADMINISTRATOR: SSA. Leo McManus and Kalman Rupp were the government project officers.

POLICY EFFECTS: Under the new "Ticket to Work" program, SSA has moved toward a system of payments to service providers that will be contingent on participant outcomes. The limited success of the cost-reimbursement based Project NetWork case management demonstration contributed to the impetus for initiating this radical change.

DATA SOURCE: Public use data are not available.

INFORMATION SOURCES: Leo McManus, Kalman Rupp, and Stephen H. Bell, “Project NetWork: A Return-to Work Project for Social Security Disability Applicants and Beneficiaries.” In: *Partners for Independence: Models that Work*, Conference Proceedings. The Fifth North American Regional Conference of Rehabilitation International, 1993;

Kalman Rupp, Stephen H. Bell, and Leo McManus, “Design of the Project NetWork Return-to-Work Experiment for Persons with Disabilities.” *Social Security Bulletin*. Vol. 57, No.2, 1994;

Michelle Wood, Valerie Leiter, Debbie M. McInnis, and Stephen H. Bell, *Case Management at Work for SSA Disability Beneficiaries: Process and In-Program Results of the Project NetWork Return-to-Work Demonstration*, Abt Associates, September 1996;

Kalman Rupp, Michelle Wood and Stephen H. Bell, “Targeting People with Severe Disabilities for Return-to-Work: The Project NetWork Demonstration Experience.” *Journal of Vocational Rehabilitation*. Vol. 12, No.1. 1996;

Valerie Leiter, Michelle L. Wood, and Stephen H. Bell, “Case Management at Work for SSA Disability Beneficiaries: Process Results of the Project NetWork Return-to-Work Demonstration.” *Social Security Bulletin*. Vol. 60 No.3, 1997;

Kalman Rupp, Dianne Driessen, Robert Kornfeld, and Michelle Wood, “The Development of the Project NetWork Administrative Records Data Base for Policy Evaluation,” *Social Security Bulletin*. Vol. 62, No.2, 1999;

Robert J. Kornfeld, Michelle L. Wood, Larry L. Orr, and David A. Long, *Impacts of the Project NetWork Demonstration: Final Report*, Abt Associates, March 22, 1999;

Robert Kornfeld and Kalman Rupp, “The Net Effects of the Project NetWork Case Management Experiment on Participant Earnings, Benefit Receipt, and Other Outcomes.” *Social Security Bulletin*. Vol. 63, No.1, 2000;

Kalman Rupp and Stephen H. Bell, “Provider Incentives and Access in the Ticket to Work Program: Implications of Simulations Based on the Project NetWork Field Experiment.” In: *Paying for Results in Vocational Rehabilitation*. (Kalman Rupp and Stephen H. Bell, Eds.) Washington, D.C.: The Urban Institute, 2003.

DAYTON WAGE-SUBSIDY VOUCHER EXPERIMENT

SUMMARY: This demonstration, conducted from 1980 to 1981, tested the effects of a tax credit voucher for employers or a direct cash subsidy to employers on a medium-sized sample of welfare recipients. Subjects were followed for one year.

TIME FRAME: December 1980—May 1981; data collected, same period; no final report exists. A Brookings Institution working version of the cited article is dated August 31, 1984.

TREATMENTS TESTED:

1. Tax credit voucher. These experimentals received a Job-Club-type treatment. They were also given both vouchers, which employers could use to obtain credits on their tax returns, and training and written materials on using this subsidy to good advantage in their job search. The credit was good for 50 percent of wages paid in the first year of employment, with the total credit not to exceed \$3,000, and 25 percent of wages in the second year of employment, with total credit not to exceed \$1,500.
2. Direct cash subsidy. These experimentals received the same treatment as the first group, except that the subsidy was to be paid quarterly directly by the program operator, without regard to the employer's tax liability.
3. Controls. Controls received the Job Club treatment, but no voucher. They received instead a one-day course in using job Bank listings at the local employment service (a placebo treatment). They were not told that they were eligible for wage subsidies (although by law they were).

OUTCOMES OF INTEREST: (1) Placement in an unsubsidized job; and (2) Initial wage.

SAMPLE SIZE: Tax credit group, 247; direct cash group, 299; controls, 262.

TARGET POPULATION: There were two: (1) Recipients of general assistance, typically single, or if a member of a couple, childless ("many were only temporarily destitute"); and (2) AFDC recipients, whose participation was not mandatory. However, many of those referred were WIN mandatory who had been unsuccessful in other WIN activities.

MAJOR FINDINGS:

1. Job placement rates. Controls, 20.6 percent; tax credit group, 13.0 percent; direct cash group, 12.7 percent. The advantage of the controls over the experimentals is statistically significant.
2. Use of the vouchers. Of 70 voucher-holders who found employment, only 19 worked for firms that requested certification for wage subsidies.
3. Initial wage rate. Average initial wages in all three groups were nearly identical.
4. Job-finding rates of AFDC and general assistance groups were almost equal. "When economically disadvantaged workers are clearly identified to potential employers as disadvantaged (by a subsidy offer), their chances of employment are harmed."

TIME TRENDS IN FINDINGS: No data available beyond eight weeks.

DESIGN ISSUES:

1. The experiment was conducted during a period of high and growing unemployment in the local area.
2. It is likely that marginally employable welfare recipients are not in a good position to explain and market their own wage subsidies. It is possible that job developers could improve on the results found here, but it is not clear that this improvement would be sufficient to justify the subsidy itself.
3. The absence of process analysis means that no documentation exists on any extraneous influences that might contaminate the findings.

GENERALIZABILITY: The experiment took place in a depressed region, and its abrupt termination leaves open some questions about its administration. However, other experiments with wage subsidies (for example, the "Wage-Subsidy Variation Experiment," and the "Florida TRADE Welfare for Work" experiment) have results that are not inconsistent with those reported here.

FUNDING SOURCE: U.S. Department of Labor, Office of the Assistant Secretary for Policy, Evaluation and Research (ASPER).

TREATMENT ADMINISTRATOR: The CETA agency in Montgomery County, Ohio. Key personnel: Not available.

POLICY EFFECTS: Targeted wage subsidies have been scaled back since study was published, but they have not disappeared. Effects on policy are unknown.

DATA SOURCES: We have no information about a public-use file for this demonstration.

INFORMATION SOURCES: Gary Burtless, "Are Targeted Wage Subsidies Harmful? Evidence from a Wage Voucher Experiment," *Industrial and Labor Relations Review*, October 1985: 105–14.

LIVING INSURANCE FOR EX-OFFENDERS (LIFE)

SUMMARY: This demonstration, conducted from 1972 to 1975, tested the effects of temporary financial assistance and job placement services on a medium-sized sample of male ex-offenders. Subjects were followed for one year.

TIME FRAME: Fiscal years 1972–74; data collected, 1972–75; final report, 1978.

TREATMENTS TESTED:

1. Controls. No treatment.
2. Financial aid. Sixty dollars a week for 13 weeks, conditional on not being reimprisoned. If the subject had earnings above \$40 a week, 50 percent of those earnings was subtracted from the weekly payment and deferred to a later week, thus slightly extending the 13-week period.
3. Job placement services. Staff members worked full time finding job openings, chauffeuring experimentals to interviews and helping them fill out job applications, and advocating on experimentals' behalf with employers and bureaucrats.
4. Financial aid and job placement. Experimentals received both job services and financial aid.

OUTCOMES OF INTEREST: (1) Rearrest; (2) Employment; and (3) Earnings.

SAMPLE SIZE: Group 1, 108; group 2, 108; group 3, 108; group 4, 108.

TARGET POPULATION: Male ex-offenders returning to Baltimore from prison, nonaddicts, with records of multiple prior offenses, at least one of them for theft; under age 45; and having less than \$400 in savings.

MAJOR FINDINGS:

1. Job placement services had no statistically significant impacts. The remaining findings are reported for financial experimentals (groups 2 and 4) versus financial controls (groups 1 and 3).

	Experimentals	Controls
New arrests, all theft crimes	48	66
Estimated new arrests, from regression with other factors	48.6	66
Estimated new arrests, from probit with other factors	45.4	66
New arrests, all crimes	107	123
In school or training, first quarter	3.7%	1.4%
In school or training, second quarter	4.2%	2.0%
Employed full-time, fourth quarter	54.7%	49%

Note: All of the differences above are statistically significant; however, schooling differentials are not significant after the second quarter, and employment differentials are not significant in the first, second, and third quarters. There are no statistically significant differences in weekly earnings in any quarter.

2. Benefit–cost analysis from a social perspective: Lower bound—\$108,565 benefits, \$27,000 costs; 4.02 benefit–cost ratio. Upper bound—\$870,431 benefits, \$16,200 costs; 53.73 benefit–cost ratio.

Benefit–cost ratios from other perspectives:

	Lower	Upper
Budgetary	0.49	2.67
Nonparticipant	0.77	3.99
Participant	1.93	3.76

TIME TRENDS IN FINDINGS: The difference in arrests was 16 in the second year, compared with 18 in the first year, indicating the effect did not disappear. However, the second-year data are lower in quality since they rely on Baltimore area court records, and some subjects had left the area.

DESIGN ISSUES: The primary grounds for uncertainty are the value of increased output and the size of the losses from theft. Benefit–cost ratios reflect various assumptions about the social discount rate, the rate of decline in the effect on recidivism, the costs of the judicial system, direct losses from theft, displacement effects, and the pattern of change in the dependence on welfare payments.

GENERALIZABILITY: The sample selection criteria were deliberately chosen to assemble a group that was likely to show a strong response. Since a policy would necessarily embrace a larger group, the “Transitional Aid Research Project” (TARP) was funded to determine whether the experimental effects would be repeated in a wider population. Lenihan’s (1978) report noted that Baltimore had fairly inexpensive inner-city housing at the time of the experiment, and thus the experimental response was obtainable there at a lower cost than in some other cities.

FUNDING SOURCE: U.S. Department of Labor, Employment and Training Administration.

TREATMENT ADMINISTRATOR: Bureau of Social Science Research.

POLICY EFFECTS: Kenneth Lenihan has stated that a California legislator, citing LIFE results, eventually obtained financial assistance for released prisoners.

DATA SOURCE: We have no information about a public-use file for this demonstration.

INFORMATION SOURCES: Charles D. Mallar and Craig Thornton, “Transitional Aid for Released Prisoners: Evidence from the LIFE Experiment,” *Journal of Human Resources*, Spring 1978: 208–36. Kenneth J. Lenihan, *Unlocking the Second Gate: The Role of Financial Assistance in Reducing Recidivism among Ex-Prisoners*, R&D Monograph 45, U.S. Department of Labor, Employment and Training Administration, 1978.

KENTUCKY WORKER PROFILING AND REEMPLOYMENT SERVICES (WPRS) EXPERIMENT

SUMMARY: This evaluation, conducted from 1994 to 1999, used a unique “tie-breaking” experimental design to estimate the impacts of the Worker Profiling and Reemployment Services system on a large sample of Unemployment Insurance claimants. Subjects were followed for two years.

TIME FRAME: Demonstration period, October 1994 – June 1996. Follow-up data were obtained for two years after the beginning of each UI claimant’s spell of unemployment. Final report, 1999.

TREATMENTS TESTED: The national Worker Profiling and Reemployment Services (WPRS) system requires Unemployment Insurance (UI) claimants who are predicted through an econometrically estimated profiling model to have a high probability of exhausting their benefit eligibility to receive employment and services early in their spell of unemployment. Members of the treatment group were notified by letter that they were required to participate in reemployment services (e.g., structured job search activities, employment counseling, and retraining) as a condition for continuing to receive benefits. Members of the control group were exempt from this requirement and, hence, did not receive the letter.

OUTCOMES OF INTEREST: Earnings, length of benefit receipt, amount of UI benefits received.

SAMPLE SIZE: Total, 1,981; treatment, 1,236; control, 745.

TARGET POPULATION: UI benefit claimants.

MAJOR FINDINGS:

1. On average, the program reduced mean weeks of UI benefit receipt by about 2.2 weeks and mean UI benefits received by \$143. Both estimates are statistically significant, although the latter at only the 10-percent level.
2. The earnings of experimentals were \$1,054 higher than those of controls. This result is statistically significant.

3. Given its low cost, the WPRS system results in cost-savings for the UI program.
4. Treatment impacts do not appear to be higher for those with higher profiling scores, suggesting that the Kentucky system of “profiling does not increase the efficiency of treatment allocation....”

TIME TRENDS IN FINDINGS: Much of the impacts resulted from an increase in early return to work and early UI exits by the treatment group relative to controls. Indeed, most of the treatment response occurred prior to members of the treatment group receiving any reemployment services. The evaluators interpret this as indicating that the program impacts mainly result from the requirement to participate in services (a deterrent effect), rather than from the actual receipt of services.

DESIGN ISSUES: Because of the experimental design, it is possible that the random assignment ratio is correlated across local UI offices with the outcome measures. The evaluators attempted to control for bias resulting from this possibility by estimating fixed effect regression models.

GENERALIZABILITY: The evaluation results should be fairly generalizable. However, because some local UI offices within Kentucky were more likely to reach capacity constraints than others, the sample is not distributed across offices in proportion to their caseload. However, the research sample had characteristics similar to those of the statewide population of UI claimants. Moreover, because of the experimental design, claimants with the highest predicted probabilities of UI exhaustion tended to be excluded from the sample. However, there was no evidence that treatment impact differed systematically by profiling scores. In generalizing to other states, it is important to recognize that UI programs and WPRS systems both vary considerably among the states.

FUNDING SOURCE: The evaluators conducted the evaluation as part of their normal academic research.

TREATMENT ADMINISTRATOR: The Kentucky Department of Employment Services.

POLICY EFFECTS: None known.

DATA SOURCE: The analysis relied on administrative data. Data available subject to privacy restrictions. Contact the evaluators.

INFORMATION SOURCES: Dan Black, Jeffrey Smith, Mark C. Berger, and Brett J. Noel, “Is the Threat of Reemployment Services More Effective Than the Services Themselves? Evidence from Random Assignment in the UI System,” *The American Economic Review*, September 2003, pp. 1313-1327.