University of Wisconsin-Madison

Institute for Research on Poverty

Discussion Papers

David Greenberg
Michael Wiseman

WHAT DID THE WORK-WELFARE DEMONSTRATIONS DO?

DP # 969-92

Institute for Research on Poverty
Discussion Paper no. 969-92

What Did the Work-Welfare Demonstrations Do?

David Greenberg
Department of Economics
University of Maryland Baltimore County
and
Institute for Research on Poverty

Michael Wiseman
Robert M. La Follette Institute of Public Affairs
University of Wisconsin-Madison
and
Institute for Research on Poverty

April 1992

The authors wish to acknowlege the substantial contribution made to the completion of this paper by the Institute. The research upon which this report is based was sponsored by the Institute for Research on Poverty at the University of Wisconsin-Madison and paid for by the U.S. Department of Health and Human Services. Michael Wiseman also gratefully acknowledges the support of the Centre for Social Policy Research at the University of Bremen, Germany. The authors are responsible for the contents of this paper and the opinions and judgments expressed herein; nothing in this paper should be construed as reflecting the positions or policy of the Centre for Social Policy, the Institute for Research on Poverty at the University of Wisconsin-Madison, or the U.S. Department of Health and Human Services. This paper was first presented April 20, 1990, at a conference on "Evaluation Design for Welfare and Training Programs" in Airlie, Virginia, sponsored by the Institute for Research on Poverty and the U.S. Department of Health and Human Services. We have received helpful suggestions and corrections from many people; we have especially benefited from careful reading of the first manuscript by Judith Gueron and Robinson Hollister.

Abstract

The 1981 Omnibus Budget Reconciliation Act (OBRA) spawned numerous welfare employment demonstration programs that sought to encourage and require adult applicants and recipients of Aid to Families with Dependent Children (AFDC) to participate in job search, training, or public employment programs. The authors review evaluations of these demonstration programs and assess the influence they have had on welfare policy and social science research. In so doing, they explain how the OBRA demonstrations worked, how they were evaluated, what impact they had on the earnings of AFDC recipients, and how successful they were in reducing the AFDC caseload. The authors pay particular attention to the efforts of the Manpower Demonstration Research Corporation (MDRC), the private nonprofit organization that evaluated many of the OBRA demonstration programs. They conclude that MDRC deserves credit for entrepreneurship in encouraging states to conduct rigorous evaluations of their innovations, for sound judgment in choice of methodology for evaluating the outcomes, and for conservatism in drawing inferences from the results. They point to opportunities for improvement in evaluation of process, impact, and benefits and costs of future demonstrations. While appreciating the historic role of MDRC, they argue for development of more systematic methods for encouraging implementation and rigorous evaluation of state welfare reforms.

What Did the Work-Welfare Demonstrations Do?

Table of Contents

What Did the Work-Welfare Demonstrations Do?*

I. Introduction: OBRA and the Demand for Welfare Innovation

A Tale of Two Decades

Here is a paradox: Welfare reform failed in the Nixon-Carter era despite ardent efforts and funding commitment by both administrations. In contrast, welfare reform succeeded in the Reagan years despite considerable executive skepticism and a burgeoning federal deficit. Here is another: At least in retrospect, it appears that social science research confounded policy-making in the 1970s but facilitated consensus building in the 1980s. The contrast between welfare policy developments during the two decades and the role of social science in each illustrate the episodic character of welfare reform, but, on a more positive note, suggest that perhaps the relation between social science and public policy may have changed. In the words of one observer, the welfare reforms accomplished under President Ronald Reagan "may come to seem historic . . . as a landmark in the changing relation between research and legislation, between knowledge and national policy-making" (Szanton, 1991, p. 591).

Consider the contrast further. From the perspective of welfare policy, the 1970s effectively began in 1969 with President Nixon's Family Assistance Plan (FAP) proposal and the first reports from what was to be a series of experiments aimed at assessing the consequences of adopting a national system of income maintenance for low-income families. The decade effectively ended with the congressional rejection of modest incremental reforms of the Aid to Families with Dependent Children (AFDC) program recommended in 1979 by the Carter administration in an attempt to save some aspects of the comprehensive Program for Better Jobs and Income (PBJI) that the president had

*This paper is an expanded version of Greenberg and Wiseman (1992), which is based upon the analysis presented herein. Some of the material presented here is incorporated in Wiseman (1991a) and Wiseman (1991b).

proposed in 1977. The clear intention of both the Nixon and Carter administrations was to enhance and consolidate the role of the federal government in providing assistance for the poor. The result was virtually nil: no significant reform was accomplished after 1972, when aid to the destitute elderly and disabled was federalized in the Supplemental Security Income program. Moreover, the investment in policy-related research seemed to add to, rather than reduce, confusion about appropriate welfare policy.[1] This confusion reflected controversy over both the likely effects upon work effort of expanding social assistance to the working poor and the extent to which providing general income support to two-parent families would promote family stability.[2]

The 1980s began with a determined attempt by the Reagan administration to reduce federal involvement in welfare and to increase the latitude granted states in operating welfare systems. Like the negative tax initiatives of the 1970s, this movement also produced a remarkable quantity of policy-related research. But the 1980s was more successful than the decade of the negative income tax: instead of a collection of dead proposals suitable only for postmortem, we had the Family Support Act of 1988, a major (if modest) national initiative passed by commanding majorities in both houses of Congress. And both observers and participants credit the research produced during the decade with significant effects on both the design and political support for this new legislation (Baum, 1991; Haskins, 1991; Szanton, 1991).

We are interested in the connection between research and policy and the lessons learned from the Reagan years for future welfare program evaluation and policy-making. To study this, we focus upon research covering the consequences of programs of employment or employability enhancement for welfare recipients that resulted from the reform initiatives of the Reagan years. We emphasize employment and employability enhancement because initiatives in this area were the core of most of the welfare innovations of the past decade.[3] More precisely, our focus is upon research covering the consequences of programs of employment or employability enhancement for welfare applicants or recipients that resulted from the Omnibus Budget Reconciliation Act of 1981 (OBRA) and subsequent OBRA-related legislation. In the remainder of this section, we set the stage by reviewing the operation of public assistance for families with children in the United States and the innovations created by OBRA.

Local Agencies in Welfare Reform

The core of the American social assistance system for families is the AFDC program. Since its inception in the Social Security Act of 1935, AFDC has been paid for by a combination of federal, state, and, in some cases, local funds and operated by the states through local bureaus that are branches of either state or county governments. An important consequence of the elevation of employment as a goal of welfare policy has been the renewed attention paid to the role of local welfare authorities in delivering social assistance and implementing welfare policy. This in itself is a fundamental reversal in the direction of welfare reform. To see why, it is useful to review a bit of the background of OBRA.

The Social Security Act grants states considerable discretion in the determination of benefits and over certain aspects of eligibility. The consequence of this discretion has been significant variation across states in the treatment of families in comparable circumstances. As a result, reformers have stressed the gains in horizontal equity that would result if the federal government assumed more responsibility for income maintenance and if tighter national standards for operations, coverage, and benefits were imposed. Such adjustments were central to FAP, to PBJI, and to the Carter proposals of 1979. This movement has met with successes in many areas, including the establishment of uniform national standards for aid to the nonelderly disabled through Supplemental Security Income and, more recently, the development of common state procedures for the assessment and procurement of support from noncustodial parents for their children. But the central goal, a uniform national income transfer system, has proved impossible to attain.

The principal roadblock to national assumption of welfare responsibilities has traditionally been cost, especially given what appears to be a political necessity to assure that reform will not harm welfare recipients in high-benefit states. But a change in the political consensus regarding the appropriate object of the welfare system has also played a role. Beginning in the 1960s, AFDC policy has moved haltingly from emphasizing cash payments to encouraging recipients to become financially independent through employment. Employment is a central concern of this national reform tradition. But the political and academic discussion of approaches to employment has evolved along two quite different paths. One has emphasized enhancement of financial incentives for work and the facilitation of responses to those incentives through the provision of support services. The other has placed greater emphasis on making work and work-related activity obligatory, again with facilitating work through the provision of support services.

The first of these approaches, incentives-with-facilitation, is clearly evident in the 1967 amendments to the Social Security Act, and it appears in FAP and PBJI. The centerpiece of the 1967 legislation was an innovation in welfare payment procedures, the "$30 and 1/3" earnings disregard, which increased the financial incentives for recipient employment. Employment was supported through the Work Incentive Program (WIN), which provided training, education, social services, and counseling. Participation in the initial WIN program was voluntary except for the unemployed fathers receiving AFDC benefits in the states that provided welfare to intact families; in practice, the WIN obligation was rarely enforced. Likewise, the Family Assistance Plan, which included a more generous income "disregard" than that provided by Congress for AFDC, incorporated only a nominal work requirement. While "Jobs" was part of the middle name of PBJI, participation in the jobs program the Carter proposal envisioned was not a requirement for receiving benefits (although grant amounts were to be reduced if work was refused). Rather, the core of PBJI was the provision of benefits through an incentive-enhancing wage subsidy. Of course, for wage subsidies to work, people must be able to obtain employment, so the wage subsidy was to be facilitated under PBJI by government guarantee of a minimum-wage job of last resort. The financial incentives proposed by FAP and PBJI were, at least in concept, uniform nationally, as was the (minimum) wage for the jobs program. This was consistent with the reform goal of a uniform national welfare system.

While initially oriented toward work facilitation, the WIN program was to become the vehicle for policy development growing out of the alternative, obligation-oriented approach to welfare employment. Making welfare recipients work for benefits has a long and frequently dishonorable history that has made policymakers suspicious of and reluctant to advance proposals tainted by requirements. But in the 1970s, talking about obligation became more respectable. This second trend was reinforced by several ideas that, by the early 1980s, were widely discussed. One of these ideas was that "comprehensive" welfare reform was an impossible goal because of the incremental character of all successful welfare policy changes (Doolittle et al., 1977). Thus, rather than clearing the slate for something "radical" like PBJI, welfare reform should begin with what the country already had: AFDC and WIN. A second idea--one more appropriately labeled a conclusion--was that for structural reasons it was impossible to rely upon work incentives contained in the welfare payment system as a means of increasing self-support among welfare recipients. In fact, it was possible that work incentives of the 1967 variety would increase, rather than reduce, welfare dependence. A third idea was that for social, political, and technical reasons, it was essential that a tighter linkage be drawn between welfare and efforts directed toward achieving self-support. The original congressional vision of AFDC was as a means of support for women and children in largely static circumstances. The new picture was dynamic: AFDC would help poor adults experiencing temporary spells of dependence who would then move on. In this new conception, welfare was still a safety net, but the emphasis was less upon softening the consequences of poverty and more upon enabling those who become poor to spring back.

The change from a static to a dynamic emphasis in welfare policy has resulted in a changed perspective on the meaning of work obligation. The traditional concept of "workfare" emphasized the obligation of working for benefits. By the 1980s, the obligation concept had been transformed from emphasis on making welfare just a low-wage job to associating public assistance with an obligation on the part of the recipient to enhance his or her employability in order to become self-supporting.

It is this tightening of the linkage between self-support efforts and public assistance that has refocused attention on the role of local authorities as agents of welfare delivery and reform. The importance of local operators to the reform movements of the 1980s is explained in part by technology, or perhaps more correctly, uncertainty about technology. When benefits eligibility criteria in welfare systems can be defined with reasonable precision, and all eligibles are entitled to assistance, local discretion in operations may serve no end. But the formulation of appropriate individual programs for employability enhancement requires consideration of many more factors than does the provision of income maintenance alone. Furthermore, any such plan must take into account labor market conditions, as well as personal characteristics, and since receipt of employability enhancement services is not as yet an entitlement, determination of services appropriate for any individual recipient is dependent upon what is to be or has been done with all other people who are eligible for the services paid for from the same budget. It may be possible for operatives in the Baltimore office of the Social Security Administration to determine if a cancer patient is eligible for Supplemental Security Income even if she lives in Tucson; the same rules apply in every location, the entitlement is open-ended, and the required information is reasonably well understood. It is far more difficult to assess how successful an able-bodied mother of three would be in a particular local training program and to weigh the gains from using program resources on her behalf against using those same resources for someone else.

The OBRA Contributions

OBRA made a major contribution toward the development of methods that integrate income maintenance with employability enhancement. Like most welfare innovations, OBRA was the product of compromise. President Reagan and his associates came to the White House with a program for welfare reform based on what they believed had been accomplished by then-Governor Reagan in California in 1971. The California reform package featured reduction of the financial incentives for employment incorporated in the calculation of benefits. But it also included a work requirement, the California Work Experience Program (CWEP), which was designed to link AFDC payments to mandatory community service without pay. The administration proposed similar innovations for the country as a whole. Congress agreed to changes in the procedures for calculating benefits, but instead of mandating work-for-welfare, OBRA provided a collection of possible building blocks or tools for an enhanced employment-oriented welfare system. CWEP (now Community, not California, Work Experience) was one of these tools.

The OBRA tool kit was expanded and modified by subsequent legislation, but the basic components have remained the same and have been carried forward as program options under the Family Support Act. These include the consolidation of authority for AFDC-related employment operations within state welfare agencies; community work experience; a work supplementation program that permitted the use of welfare grants to subsidize client employment in public, private nonprofit, and (after 1984) private for-profit organizations and businesses; and expanded authority for requiring participation in job search programs by both AFDC applicants and recipients.

Two features of the OBRA tool kit, one administrative and the other operational, are of particular importance. On the administrative side, the WIN demonstration authority signaled the beginning of the end of a two-prong, two-agency welfare operations strategy in which one set of organizations--the U.S. Department of Health and Human Services and state welfare agencies--were responsible for income maintenance and another--the U.S. Department of Labor and state employment agencies--were responsible for administration of job search and training requirements. This structure was established in 1967 with the original authorization of the Work Incentive Program and had become, in the opinion of many critics, a barrier to the integration of employment preparation and income maintenance policy. This obstacle was more than bureaucratic: it was spatial. In many communities referral of eligible welfare applicants and clients to WIN services literally included a map to the location of a different agency. Separation of powers and responsibilities inevitably created communication problems and, possibly, lack of association for participants between receipt of income from one source--welfare--and the obligation to utilize the services of another--the state employment agency. OBRA's WIN-Demonstration authorization offered states the opportunity to put overall responsibility for WIN/welfare integration in the hands of the public assistance agency and led to literal collocation of welfare intake/benefit determination processes with services related to employment.

The multiplicity of tools in the OBRA kit has important consequences for welfare employment program operations. It is not possible to assign the same recipient, at the same time, to all of the activities encompassed by the legislation. As a result, it is necessary to develop procedures for sequential assignment of clients to various activities and monitoring their progress through this sequence. From the beginning, the Social Security Act provisions that authorize AFDC have given states options for setting levels of benefits and for limited variation in program coverage. OBRA and its successors provided greater latitude to the states in choice of employment-related activities. The states' response to the opportunities provided by the new legislation was to generate a series of programs constructed from these elements. Although interesting as an example of creative federalism, the outpouring of innovation from the "thousand flowers" welfare policy encouraged by OBRA was potentially a disappointment for those interested in systematic improvement of policy. To identify innovations worthy of evaluation, we need to know what the consequences of such policies are for the behavior of families and individuals at risk of poverty and welfare dependence. While a multiplicity of approaches potentially provides many degrees of freedom for assessment of the consequences of the various OBRA-related tools and the processes created to manage them, a priori there was no reason to believe that a sufficient quantity of reliable data would be produced to allow the potential information gains from so many experiments to be realized. In fact, we have learned a great deal.

The OBRA Evaluations

The Social Security Act allows considerable variation from state to state in certain dimensions of AFDC operations, but within state borders the program must be operated in a uniform manner. Operation of OBRA-based programs required in many instances that the states be granted waivers of this uniformity requirement, since operations typically could not be implemented statewide, at least initially. Granting of waivers requires an evaluation plan. Thus, one result of OBRA-based innovations has been a substantial number of evaluations of the consequences of such activities.

The quality of many of these OBRA evaluations is very high. There are at least two reasons for this, one historical, the other entrepreneurial. While OBRA reformers were in part motivated by dissatisfaction with top-down approaches to welfare reform inherited from a decade of Washington-spawned welfare reform proposals, they were heirs as well to the results of fifteen years of welfare experimentation. Much of this work--most notably the four income maintenance experiments (Munnell, 1986) and the Supported Work Demonstrations--was based on sophisticated research studies that compared outcomes for persons (experimentals) assigned at random to receive some reform-related "treatment" to outcomes for individuals who were not eligible to receive the treatment (controls). The legacy of this tradition has been a growing academic consensus that random assignment is a methodologically superior approach to program evaluation and evidence that such studies can in fact be carried out.

No matter how strong academic support for impact evaluation on the basis of random assignment of participants to experimental and control groups might have been, it is unlikely that states would have, on their own, adopted this approach. The translation of consensus into action is attributable to the aggressive entrepreneurship of the Manpower Demonstration Research Corporation (MDRC). As architect of the Supported Work Demonstrations, MDRC was committed to random assignment as the foundation of evaluation of employment policy innovations. The firm organized private grants for the OBRA evaluation effort and offered states subsidized research in return for a state commitment to an evaluation based on random assignment of clients to control and treatment groups. In some instances, states were able to use foundation contributions as part of the local contribution qualifying for federal matching funds under AFDC funding arrangements. Once these techniques caught on, and MDRC delivered the research reports that became required reading for welfare administrators nationwide, the number of MDRC-evaluated experiments grew rapidly.

The upshot is that we began the last decade of the twentieth century with a remarkable collection of new information on the consequences of employment-related welfare innovations, much of which has been produced in a common and comparable format. This paper surveys twenty-four evaluations of individual OBRA-related state demonstration programs, thirteen of which were conducted by MDRC. In addition, we have reviewed a large collection of related research, both cross-cutting and targeted at complementary, but not specifically OBRA-based policies. Our objective is to assess what has been learned from this work about the efficacy of both the employment-related interventions themselves and the institutional and methodological procedures that have been used for their evaluation.

We begin in the next section with an overview of issues in demonstration design and evaluation. We then turn, in Section III, to a study of the welfare processes created by the OBRA demonstrations. Section IV reviews what is known about the impact of OBRA-related innovations, and Section V discusses problems that arise in conducting cost-benefit studies of these outcomes. We conclude in Section VI with an overview of what appear to us to be the national consequences of the decade of experimentation we have observed, and some reflections on the lessons learned and the work still needed.

The Family Support Act included a requirement that the JOBS program be evaluated using "experimental and control groups that are composed of a random sample of participants in the program . . ." In 1989 the U.S. Department of Health and Human Services awarded a contract for carrying out the required evaluation to the Manpower Demonstration Research Corporation. The contract called for completion of a survey of JOBS-related research results. The result is an important book, From Welfare to Work (Gueron and Pauly, 1991), that is comparable in some respects to the work reported here. However, the emphasis in From Welfare to Work is upon synthesis of a broad range of experimental results relevant to JOBS as a guide for both program operators and planning of the JOBS evaluation. The emphasis of this work, which predates the MDRC study, is specifically upon OBRA and the effectiveness of statutory mechanisms for encouraging research on the effects of innovations in welfare-related employment and training programs. We attempt wherever possible to refer to related materials in the MDRC study, but it is not our intention to conduct a detailed evaluation of it. Even so, the conclusion to this paper does summarize the approach adopted by MDRC for the JOBS evaluation.

II. The Welfare Employment Process

A Welfare Employment Program

This section sets out an organization scheme for the OBRA studies and presents an overview of the reports we reviewed. We refer to all the programs developed in response to OBRA opportunities as Welfare Employment Programs, or WEPs. Figure 1 is a flow diagram for a stylized WEP. Diagrams similar to this one are a common feature of the evaluation reports we have reviewed. The diagram depicts the sequence of activities a client experiences, with flow from top to bottom of the chart. WEPs vary considerably from one to another, and no actual program flow diagram matches Figure 1 exactly.

To depict the client's experience completely, a second diagram would be needed for the income maintenance system. This diagram would identify the sequence of procedures used for determining welfare eligibility, assessing financial need, and mailing checks. At some point in the process of applying for income assistance, or at some point after eligibility is established, adults are "referred" to the WEP. The referral is triggered by a set of eligibility criteria. Note that while the welfare process is case or family oriented, the WEP process is organized around individuals. Since not all welfare cases include WEP eligibles (often considerably less than half do), and since WEP participation may begin before family AFDC eligibility is established (indeed, in some systems, participation of eligibles is a precondition for acceptance of the family for AFDC), not all WEP client families have yet been accepted as welfare cases, so there is not a one-to-one correspondence between the records of the two systems.

In the program represented in the diagram, WEP registration is followed by assessment of immediate service needs--child care, for example. Once such needs are met, the participant begins a period of job search. If job search fails to produce employment, a more extensive case review is

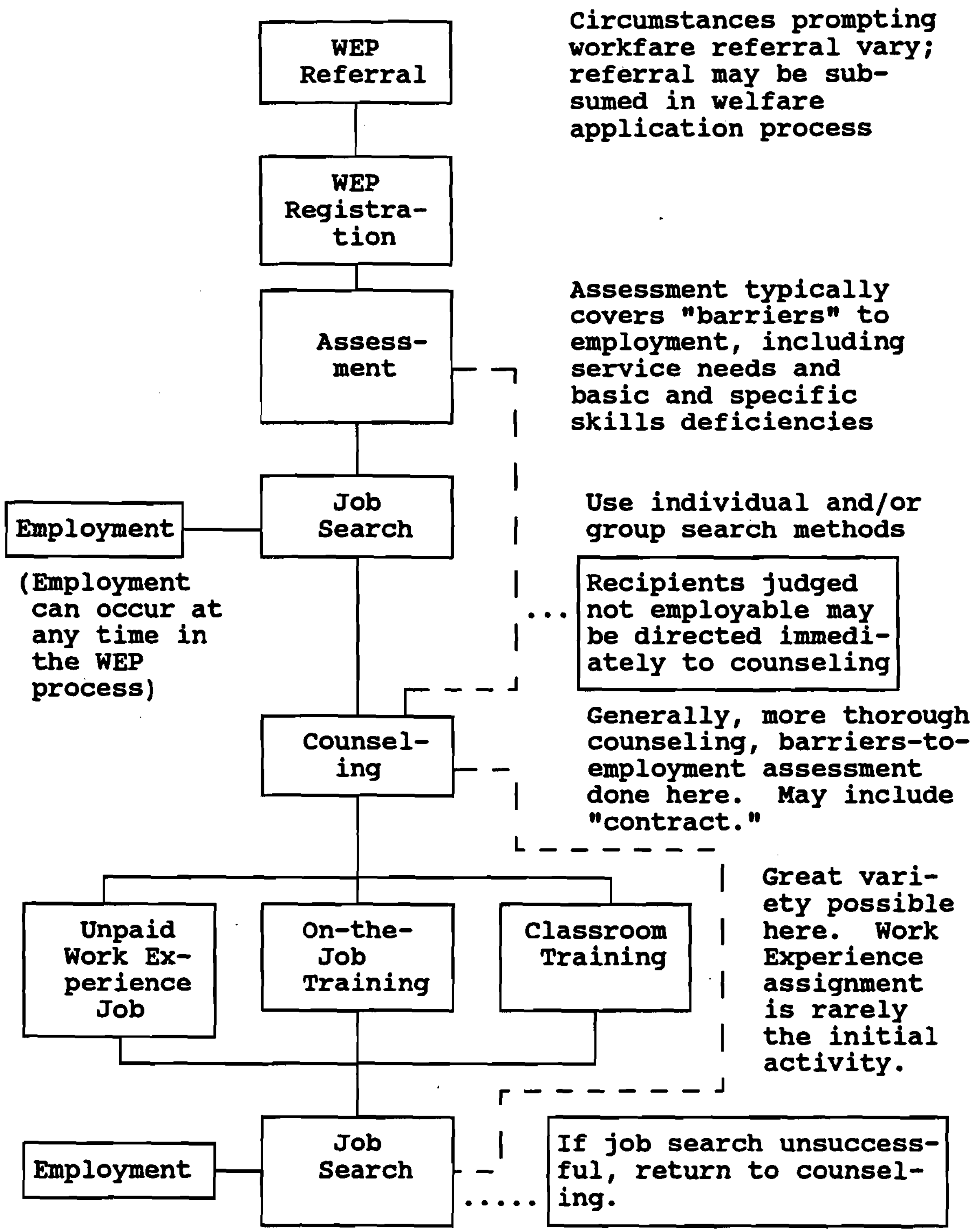

Figure 1: Flow Chart, Hypothetical Welfare Employment Program

done, possibly with a "contract" for future employment-related activities as the product. The next steps in this relatively elaborate program involve participation in those activities. Note that some clients bypass initial job search and receive counseling instead. Of course, the amount of structure imposed on the activities, the amount of oversight ("case management") exercised, and the predictability of the experience of participants in the system vary from program to program.

Figure 1 illustrates three properties shared by many of the programs we reviewed. First, this WEP is a process rather than a condition. Traditionally, saying that a client was "on welfare" sufficed as a description of his or her condition because welfare was a stationary state. In the past most case management information systems used by welfare agencies maintained little or no historical information on case status since, in principle, eligibility was determined solely by a family's current status. But welfare employment processes are time-dependent, and as a result, "in WEP" leaves something more to be described. Second, most of these programs may be divided, like traditional casework, into a series of steps, or components, from which the process is constructed. It is often the case that components are provided by other agencies or nonprofit organizations. Indeed, in some WEPs, even the process is managed by an agency other than the income maintenance unit. Finally, the process may be tailored to individual characteristics, and this tailoring will show up in variation in the sequence of steps recipients take. To the extent the differences in these sequences are systematic, they may be termed "tracks."

Objects of Interest for WEP Evaluation

A WEP evaluation may be focused on any or all of several levels. In the organizational sense, the highest level is the political/institutional process that creates such a system. OBRA and concurrent budgeting decisions basically presented states with an opportunity and a change in resources. States responded to this offer with programs that mixed the opportunities in varying ways. A political/institutional process analysis would presumably look for federal and state motivations and attempt to identify congruencies and conflicts in these motivations and the consequences of such factors for the program. Which objectives were realized, why, and to what extent?

We have not included the political process in our overview in any systematic way. But we draw attention to it to emphasize that OBRA provided more than one type of demonstration. At the national level, the OBRA experiment could be interpreted as an effort to see what sort of responses the set of opportunities might engender; in principle, these responses could be studied in essentially the same way as, on the individual level, we study the responses of households and firms to fiscal incentives in, for example, the tax code. But it is important to note that we do not have comparable information on all state responses. We have in hand only evidence on outcomes for states that applied for waivers of certain Social Security Act provisions for WEP operation and/or were induced to submit their programs to the scrutiny of MDRC or other evaluators. If the jurisdictions that undertook evaluations of OBRA-based experiments are exceptional--and we know they are--then making inferences for the consequences of similar programs for the nation as a whole from the results of the experiments is hazardous at best. The odds that three draws in a random sample of twenty-four OBRA-related demonstrations nationwide would come from San Diego County--as is true for our survey--are extremely low!

The hazard of selection bias is reduced if we can uncover the "recipes" that are at the heart of the various WEPs created as a result of the OBRA opportunity. This is the second possible evaluation focus, and it can be approached in two ways. One is to emphasize structure, including the range of contributing organizations, bargains struck between the managing agency and the contributors, effects of the innovation on the sponsoring organizations, and public presentation of the program. The alternative is to emphasize the experience of individual WEP participants, that is, to describe the WEP "treatment." Identifying the treatment is essential if effects are to be replicated. Both the investigations of organization and treatment have been termed "process analysis."

A third evaluation emphasis is upon the effect of the recipe upon client behaviors and outcomes. Known as "impact analyses," the outcomes of interest have traditionally included employment, earnings, and welfare utilization. While such evaluations may concentrate on broad-based comparisons of the effects of one general system to another (or perhaps to no system at all), once a WEP is in place, the decisions that count from the local agency perspective are those that marginally change program operation--for example substituting one method of job search for another.

Once a substantial number of evaluations of various types are in hand, a fourth object of evaluation can be the evaluations themselves: What policies were tested? Why? What designs were utilized?

Finally, for policy analysis, the object is often some sort of cost-benefit analysis in which program outcomes are summarized to whatever extent possible in dollar terms. This common unit allows assessment of the efficiency of the project--whether the value of resources gained exceeds resources lost.

Our interest in this paper concerns lessons for policy evaluation and insights into the consequences of WEPs for clients. Accordingly, we skip political and organizational analysis and focus instead upon review of WEP client processes, demonstration research designs, client impacts, and cost-benefit analyses. While useful political and organizational analyses of individual OBRA-related programs are available (cf. Behn, 1989), a comparative political/organizational analysis of state programs is, to our knowledge, not yet available.

The Reports

Our sample was defined solely on the basis of availability of data to us. We began by asking the Office of the Assistant Secretary for Planning and Evaluation in the U.S. Department of Health and Human Services (HHS) for all OBRA-related demonstration evaluation reports submitted to HHS through December 31, 1989, and we added all of the reports produced by the MDRC Work/Welfare

Initiatives project. We did not pursue information on demonstrations authorized but never reported on, although review of such demonstrations would be a useful part of a comprehensive study of state responses and the procedures followed by HHS in regulating state demonstration efforts. In two cases (Maryland and West Virginia), one report covered two quite different demonstrations, and we have treated them separately.

All told we accumulated reports on twenty-four demonstrations. These are summarized in Table 1, at the end of this section. We have concentrated upon twelve features, the first two of which identify the location of the project and the source of our information:

Name and Location. The OBRA demonstrations are each associated with a specific state. However, while some of the OBRA demonstrations were conducted over the entire state, others were limited to only subareas such as counties or welfare office catchments. In each case, we have listed a short form of the project designation (e.g., California San Diego SWIM) and, typically, the words behind the project acronym. Nineteen states are represented, and this variety is reflected in substantial program diversity and considerable variation in the economic environments in which innovation was carried out.

Evaluator/Reference. We report the principal (generally the latest) reference for program information. Full citations and related readings appear in the bibliography. We also indicate the organization responsible for conducting each of the listed demonstrations. As mentioned earlier, thirteen of the twenty-four evaluations were conducted by MDRC. The remaining evaluations were either conducted in-house by the state or by research firms other than MDRC.

The next two categories describe the pool of eligible participants:

Household Type. The demonstrations differ in program focus and definition of the adults targeted for participation. We use the designation AFDC-R for what is often termed the "basic" or "regular" AFDC program: income maintenance for families with children in which deprivation results from the absence or physical incapacity of a parent. AFDC-U refers to the optional (for states, prior to the Family Support Act) program for families with children made needy by the involuntary unemployment of a parent.

Target Clients. Within the AFDC-R and AFDC-U categories, the programs studied were generally directed toward adults whose participation in the Work Incentive Program was mandatory. For AFDC-R, the most significant group typically exempted from WIN participation was single parents with children younger than six.[4] As the table indicates, in five of the programs, mandatory participation was extended to parents with no children younger than three. In some demonstrations, clients who were not required to participate were allowed or encouraged to do so.

The programs vary substantially in eligibility restrictions based on welfare status at point of referral. Even when general eligibility for a particular program is defined broadly, impact analysis may be confined to a particular status subgroup. Considered in relation to the income maintenance system, four types of WEP participants can be identified. In 1972, Congress made registration for WIN an eligibility requirement for AFDC application approval. For some WEPs, and in some WEP evaluation schemes, it is welfare application that initiates WEP treatment and marks the point at which experimental groups are defined and differentiated from clients who are to serve as controls. In Table 1, "APPs" identifies this target client pool. Alternatively, the WEP may be introduced (and/or experimental participants designated) at the point welfare eligibility is formally obtained. This would be "New Recipients." A third port of entry occurs when circumstances change and recipients who formerly were not mandatory participants become so. These are designated in the table as "NDMRs", for "Newly Determined Mandatory Recipients." An NDMR is created, for example, when the youngest child in the household of a woman who otherwise satisfies the criteria for WIN participation reaches the age of six (or, in some WEPs, three). Finally, some WEPs served all AFDC recipients who were eligible for WIN at the time the program was initially established, and evaluation groups were drawn in some manner from this pool. In the table, the target client pool for this group is listed as "CMRs", for "Continuing Mandatory Recipients." Typically, a CMR might be referred to a WEP after completion of a periodic eligibility review.

The distinctions among the various classes of potential WEP participants may appear a bit arcane, but such distinctions have at least three important consequences. First, applicants are not recipients, and some WEP proponents believe early interception of persons applying for public assistance will dissuade some from taking up welfare at all. Programs pushing the intervention up to the applicant pool could possibly be analyzed to detect this effect, but this is not explicitly attempted in any of the studies we consider. Second, if programs are intended to achieve a high rate of participation of a target group of eligible clients in employment-related activities, it may initially be necessary to restrict the rate of intake of eligibles in some way in order to avoid swamping the system. Taking new mandatories--either APPs, New Recipients, or NDMRs--has the advantage of allowing the agency to deal with a flow of clients that is smaller, presumably, than the stock of CMRs. Finally, as is well understood, the welfare process "sorts" cases. On the whole, adults in the existing AFDC caseload are found to be more disadvantaged and more isolated from the labor market than adults in the applicant families. Differences in outcomes may reflect this.

We summarize the programs themselves with reference to components, sequence, obligation, and operating agency:

Components. While references to many different types of services appear in the descriptions of these programs, most individual programs are constructed from only a few components. The most common are varieties of job search, a carryover from WIN. The most significant distinction between job search types is that between directed individual search (JS-I) and search conducted in the context of group meetings (JS-G). Other services fall more or less into four categories: (1) adult basic or general education (AB-GED), (2) community work experience (CWEP), (3) various forms of skill-specific training generally done in classroom settings (marked "Training" in the table), and (4) on-the-job training (OJT) in which skills are acquired in the context of regular employment. States were permitted to require job search by welfare applicants even before their eligibility was established and payments were initiated. Exercise of this option is identified as AJS (applicant job search) in the table.

We have separately identified CWEP as authorized by OBRA and an alternative form of work experience authorized through WIN--WIN Work Experience (WWE). This distinction has important operational consequences. Under the original OBRA authorization, clients could be required to work a fixed number of hours, an amount defined by dividing the AFDC grant by the minimum wage. In principle, the client could be obligated to do this indefinitely. The purest version of such a program was offered in West Virginia. Under WIN Work Experience (the approach that has been emphasized in various demonstrations in California), the number of hours an eligible recipient could be required to work was not restricted, but the obligation could endure for no more than thirteen weeks.

Sequence. Distinctions may be drawn among WEPs on the basis of the nature of and importance attached to the sequence of activities. Programs generally fall between two extremes. In one, all participants go through the same steps. In the other, virtually nothing is prespecified. For example, the San Diego Employment Preparation Program/Experimental Work Experience Program (EPP/EWEP) had a fixed participation sequence--job search followed by CWEP. The Baltimore Options program, which utilized a variety of different components, did not have a fixed sequence but instead tried to allow caseworkers to tailor services to individual needs. Some programs, such as California's Greater Avenues for Independence (GAIN) effort, begin with a relatively rigid sequence and then at later stages in the process provide considerable caseworker and client discretion.

Requirement. In principle, all programs listed in the table could be viewed as "mandatory," since during the interval spanned by these demonstrations WIN registration was mandatory for certain welfare clients by federal law. For some, however, participation in activities beyond WIN registration was treated in practice as voluntary. Indeed, Massachusetts emphasized this aspect of its ET-Choices program, allowing AFDC recipients to decide for themselves whether or not to participate at all in the program and, if they did decide to participate, to select among a menu of available program components. Aside from this highly publicized exception, most of the OBRA demonstration programs were at least nominally mandatory in all phases. Except in GAIN, failure to comply with program requirements was supposed to be met with a "sanction": elimination of the noncompliant adult from the household's grant computation in the case of AFDC-R and loss of welfare eligibility altogether for AFDC-U cases. In GAIN, California has experimented with reduction of client expenditure discretion--third-party management of household expenditures--as an alternative to deregistration and grant reduction.

Operating Agency. While the opportunity for assumption of welfare employment program operations by state welfare agencies was viewed by many as a significant aspect of OBRA, not all states felt the option important. In Arizona, for example, both the employment and welfare departments were already part of a single, "super" agency; in some other states (for example, California and Washington), the dual-agency structure was maintained. For those WEPs operated by welfare departments, the table distinguishes between those states in which AFDC delivery is a function of county welfare departments (CWD) and those states in which local welfare offices are operated by the state welfare agency (SWA).

The remainder of the table dates the demonstrations and outlines the format of the available information.

Date. Table 1 presents an initiation date and a "study intake" time period for each evaluated demonstration. The initiation date marks the point at which each evaluated WEP began operation. The study intake time period identifies the period during which clients on whom the evaluation was based entered the program. For studies not based on analysis of individual client data (the O'Neill Massachusetts study, for example), this interval identifies the period of program operation covered by whatever technique was employed. In general, studies in which operations were conducted for a considerable period of time prior to evaluation offer the best prospect of measurement of effects independent of normal implementation problems. In some cases (for example the San Diego Saturation Work Initiative Model [SWIM] demonstration), the program evaluated was a natural extension of programs already in place and, as a result, little disruption occurred as it was phased in.

Process Analysis. Most of the reports summarized here include information on program structure and client experience. We have attempted to use such process information to identify the consequences of the innovation for client experience and, in particular, for services clients were provided. In Section III, we develop summary measures of process based on the incidence of activity among participants. The major distinction drawn is between two types of studies. One type permits measurement of the cumulative incidence of various activities among recipients over intervals measured relative to the point at which they entered the program; the other permits measurement of the incidence of program activity within the set of all eligibles at a particular point in time. These measures, respectively labeled CI (for cumulative incidence) and PITI (for point-in-time incidence) in the table, are described and illustrated later in the paper.

In assessing process outcome, as well as assessing impacts, it is important to identify the comparison group. For most, but not all, of these OBRA demonstrations, the effect of the innovation on welfare process was identified through comparing the experience of the treatment group with that of the control group. One alternative to this approach was comparisons with experience under WIN.

Impact Analysis. Section IV of this paper is devoted to an examination of the impact of the OBRA demonstrations on client employment, earnings, and welfare utilization. In general, these impacts are estimated by reference to a comparison group. The table identifies those demonstrations for which impacts were assessed, the basis for establishing the comparison group, the services that the comparison group received, and the maximum duration of program and postprogram experience over which experimental/control comparisons were conducted.

Cost-Benefit Analysis. Because both the costs and benefits of welfare innovations are varied and accrue over time, a framework for intertemporal aggregation must be used to assess the economic efficiency of these innovations. Such a framework is provided by cost-benefit analysis. As Table 1 indicates, cost-benefit analyses were conducted of a number of the OBRA demonstrations. However, these analyses raise many issues. Section V of this paper examines these issues and describes results from the analyses. We distinguish among several viewpoints or perspectives from which the benefits and costs of the OBRA demonstrations have been assessed--that of society as a whole, that of the client, that of taxpayers, and that of the government or "agency." The perspectives used in the individual evaluations are indicated in the last row of the table. As can be seen, some of the evaluations examined costs and benefits from only an agency perspective--that is, only in terms of effects on government budgets. In Section V, we argue that this perspective is far too narrow.

Five Observations Concerning the OBRA Demonstrations

Before turning to the process analysis, we make five general observations on the overview presented by Table 1. The first is that the dimensions of latitude apparent in the design, implementation, and evaluation of WEPs are substantial. As a result, we can anticipate that, even given twenty-four to work with, the likelihood of identifying clear-cut relations between outcomes and inputs adequate to serve as a basis for forecasting the consequences of national replication of such innovations is small.

A second observation is that a broad distinction can be made between two types of demonstrations: those that concentrate on implementation of an alternative to the minimal services provided by WIN, and those that are designed to provide evidence on the impact of alternative tracks within a richer WEP process. Some do both. We count as tests of components the EPP/EWEP project in San Diego (the effectiveness of adding CWEP), Florida's TRADE program, the Illinois WIN Demonstration Program (WDP), Maine's TOPS, the New Jersey OJT demonstration, the Virginia Employment Services program, and one of the two Washington state experiments, although, as discussed later, not all of these experiments yielded useful information.

A third point concerns timing. The earliest intake of client data for evaluation of one of the OBRA demonstrations began in June of 1982 (the Washington Applicant Employment Services [AES] demonstration); the last data on impacts we have utilized was collected for the third quarter of 1988 (12-quarter observations from the San Diego SWIM program). Many changes occurred in the economy during this interval. Overall, conditions improved markedly, with unemployment rates falling from an average of 9.2 percent in 1982 to 5.5 percent in 1988. At the same time, while the rate of inflation was much lower than the experience of the 1970s, prices and wages did go up. While labor markets may have tightened, in many sites, notably in California but elsewhere as well, employment opportunities for public assistance recipients were impacted by growing competition from immigrants, both documented and undocumented. We have not attempted adjustment in this paper for price effects or place-to-place and time-to-time variation in labor market conditions in this paper.

A fourth point is that these projects have as a useful by-product important information heretofore unavailable on the experience of households receiving public assistance. In particular, the OBRA demonstrations provide the richest data available on the experience of welfare-dependent families over time. For some purposes, these data might be superior to information available from national surveys such as the Panel Study of Income Dynamics or the Survey of Income and Program Participation because they are reported for quarterly and, in some cases, monthly intervals. Moreover, they are possibly more reliable than survey-provided data, at least with respect to transfers and participation in the above-ground economy, because much of the information is taken directly from institutional sources. Of particular importance is the fact that the OBRA demonstrations provide an unusually rich body of information on AFDC-U families since virtually every AFDC-U family included a WIN mandatory adult. The AFDC-U program has historically been small, both because only about half of all states had it and because, even in those states offering the option, AFDC-U caseloads were rarely more than a tenth the size of those for AFDC-R. As a consequence, any random sample of welfare recipients included very few AFDC-U observations, often too few to support reliable inferences about their characteristics. Nonetheless, intact families have long been on the frontier of income maintenance expansion, and, as a result, particular interest is attached to the consequence of OBRA-type intervention for this group.[5]

The final point is that findings from the OBRA demonstrations are dominated by the analyses of the Manpower Demonstration Research Corporation and the techniques MDRC has adopted. As a result, an important objective of this evaluation must be to assess the degree to which the MDRC approach served public ends.

Summary Overview: The OBRA Demonstrations

Pages 27-33 are taken up by Table 1, which summarizes the twenty-four OBRA demonstrations studied in connection with this review. The table is divided into three segments, and each segment covers eight demonstrations. The demonstrations are listed left to right and characteristics summarized top to bottom. The demonstration names appear at the top of each page. The characteristics categories are listed in the left-hand column of each page. Although most acronyms used in the cells are explained in detail in the text, for the reader's convenience we list what they stand for below.

AB-GED	adult basic or general education
AJS	applicant job search
CI	cumulative incidence

CWD	county welfare department
CWEP	community work experience
ESA	employment services agency
JS-G	group job search
JS-I	individual job search
JTPA	Job Training Partnership Act
OJT	on-the-job training
PITI	point-in-time incidence
PVT	prevocational training
SWA	state welfare agency
WWE	WIN Work Experience

TABLE 1

Summary Overview: The OBRA Demonstrations
(Part A: Arizona to Maine)

	Arizona WID	Arkansas WORK	California San Diego EPP/EWEP	California San Diego SWIM	California GAIN	Florida TRADE	Illinois WDP	Maine TOPS
Extended Name	Work Incentive Demonstration Program	WORK Program	Employment Preparation/ Experimental Work Experience Program	Saturation Work Initiative Model	California Greater Avenues for Independence	Welfare for Work Program	WIN Demonstration Program	Training Opportunities in the Private Sector
Location	Maricopa, Pima counties	South Pulaski, Jefferson counties	Countywide	2 of 7 county welfare offices	8 counties, 7 for participant process analysis	Most of state	Cook County	Statewide
Evaluator/Reference	MDRC:Sherwood (1984)	MDRC:Fried-lander et al. (1985)	MDRC:Goldman et al. (1986)	MDRC:Hamilton and Friedlander (1989)	MDRC:Riccio et al. (1989)	Florida DHRS (1987)	MDRC:Fried-lander et al. (1987)	MDRC:Auspos et al. (1988)
Eligible Household Types	AFDC-R	AFDC-R	AFDC-R,U	AFDC-R,U	AFDC-R,U	AFDC-R	AFDC-R	AFDC-R
Target Clients	WIN eligible recipients w/o children <3	APPs, NDMRs w/o children <3	APPs	APPs, NDMRs, CMRs	APPs, NDMRs, CMRs, volunteers	WIN mandatory recipients after WIN job search; volunteers	Welfare-approved APPs, NDMRs	Female AFDC-R recips. with >6 mos. AFDC receipt, "employable"
Program Components	JS-I, JS-G, training, job development	JS-G, JS-I, WWE	AJS, JS-G, WWE	JS-G, WWE, AB-GED, JS-I, training	AB-GED, CT, OJT, WWE, JS-G, JS-I	Two levels of wage subsidy	JS-I, WWE	"Prevoca-tional training," WWE, OJT
Sequence	Sequence not emphasized	JS-G/JS-I/WWE	Two tracks: AJS/JS-G and AJS/JS-G/CWEP	JS-G/CWEP followed by other activities	Complex tracks with AB-GED, JS-G, training, continuing services	JS-G/JS-I/OJT; test focused on effect of alterna-tive wage subsidies	Two tracks: one is JS-I, the other, JS-I/ CWEP	PVT/WWE/ OJT

(table continues)

TABLE 1
Part A (continued)

	Arizona WID	Arkansas WORK	California San Diego EPP/EWEP	California San Diego SWIM	California GAIN	Florida TRADE	Illinois WDP	Maine TOPS
Requirement	Job search mandatory	Mandatory	Mandatory	Mandatory; modified WIN sanctions	Mandatory; modifies WIN sanctions, incl. money management	Job search sequence mandatory; OJT voluntary in practice	Mandatory	Voluntary
Operating Agency	Employment services agency closely linked to welfare agency	Welfare agency (WIN Demo)	Joint state ESA and CWD	Joint state ESA and CWD	Joint state ESA and CWD	SWA with JTPA agencies	SWA	SWA with JTPA agencies
Date of Program Initiation	6/82	10/82	8/82	7/85	7/86	7/84	1/84	10/83
Date of Study Intake	10/83-7/84	6/83-3/84	10/82-8/83	7/85-6/86	12/86-6/87	7/84-9/86	2/85-9/85	10/83-12/84
Nature of Client Process Analysis	PITI	CI	CI	CI, PITI	CI	Descriptive	CI	CI
Nature of Comparison Group	Previous WIN	Impact controls	Impact controls	None	None (analysis incomplete)	None	Impact controls	Impact controls
Basis of Impact Analysis	None conducted	Random assignment	Random assignment at application	Random assignment at application or reevaluation	Not available (analysis incomplete)	Random assignment to control, two wage subsidy groups	Random assignment after welfare acceptance for APPs, status change for NDMRs	Random assignment after intensive recruitment
Comparison Group	Not applicable	No services	Minimal WIN services	Minimal WIN services	Not available	Regular WIN services	Minimal WIN services	Regular services without TOPS sequence
Maximum Duration of Observations	Not applicable	12 qtrs.	6 to 9 qtrs.	9 to 12 qtrs.	2 (process study)	2 qtrs.	6 to 8 qtrs.	12 to 16 qtrs.

(table continues)

TABLE 1
Part A (continued)

	Arizona WID	Arkansas WORK	California San Diego EPP/EWEP	California San Diego SWIM	California GAIN	Florida TRADE	Illinois WDP	Maine TOPS
Benefit-Cost Analysis Perspective	None conducted	Client, agency, taxpayer, social	Client, agency, taxpayer, social	Client, agency, taxpayer, social	Not available (analysis incomplete)	Agency costs/ benefits	Client, agency, taxpayer, social	Client, agency, taxpayer, social
Comments	Principally organizational study	Exceptional impact for few resources	Pioneer work/ welfare experiment	Emphasis on client process, participation rates; large impacts	Available reports focus on organization effects, participation	Administrative failure	Minimal impact for few resources	Small scale, selective; large earnings impacts for those selected

TABLE 1

Summary Overview: The OBRA Demonstrations
(Part B: Maryland to Oregon)

	Maryland Baltimore Options	Maryland BET	Massachusetts ET-Choices	Minnesota CWEP	New Jersey Grant Diversion	North Carolina CWEP	Ohio CWEP	Oregon JOBS
Extended Name		Basic Employment and Training	Employment and Training Choices					
Location	Baltimore City	Wicomico County	Statewide	8 counties	9 counties	6 counties	5 counties	Statewide
Evaluator/Reference	MDRC: Friedlander (1987)	MDRC: Quint et al. (1984)	Massachusetts DPW (1986); O'Neill (1990)	Minnesota DHS (1987)	MDRC: Freedman et al. (1988)	North Carolina DHR (1985)	Potomac Institute (1988)	Oregon DHR (1988)
Eligible Household Types	AFDC-R,U	AFDC-R,U	AFDC-R,U	AFDC-U	AFDC-R	AFDC-R	AFDC-R,U	AFDC-R,U
Target Clients	APPs, NDMRs	APPs, NDMRs, CMRs, volunteers	All adult welfare recipients	All adults not specifically exempt	All adults deemed "employable"	APPs, WIN eligible recipients w/o children <3, volunteers	All WIN eligible recipients	APPs w/o children <3
Program Components	Variety of JS, training, AB-GED, WWE, CWEP, OJT	JS-G, AB-GED, training, OJT, WWE	Variety of training placement, OJT, AB-GED activities, child care support emphasized	JS-I, JS-G, CWEP	Regular WIN services, OJT	JS-G, employment preparation, CWEP	JS, CWEP, OJT, supported work	AJS, JS-G
Sequence	Individualized sequence	JS-G followed by choice of paths	Individualized sequence	JS, followed by CWEP	WIN services of any kind/OJT placement	Employment preparation/JS-G/ CWEP	Information not available	AJS/JS-G
Requirement	Mandatory "when judged appropriate"	JS-G mandatory	Voluntary	Mandatory	Voluntary	Mandatory	Mandatory	Mandatory
Operating Agency	SWA with Baltimore JTPA agency	SWA	SWA	CWD	CWD	CWD	CWD	SWA
Date of Program Initiation	11/82	10/82	10/83	Spring 1983	4/84	7/82	2/83	1/82

(table continues)

TABLE 1

Part B (continued)

	Maryland Baltimore Options	Maryland BET	Massachusetts ET-Choices	Minnesota CWEP	New Jersey Grant Diversion	North Carolina CWEP	Ohio CWEP	Oregon JOBS
Date of Study Intake	11/82-12/83	10/82-12/83	1983:4-1987:4	10/84-9/86	10/84-9/85	Not available	1981:1-1987:3	Not applicable
Nature of Client Process Analysis	CI	CI	Descriptive	PITI (average)	CI	PITI	None	None
Nature of Comparison Group	Impact controls	None	Pre ET-Choices WIN	None	Impact controls	None	Not applicable	Not applicable
Basis of Impact Analysis	Random assignment at application for APPs; at redetermination for NDMRs	None conducted	Time series analysis; interstate comparison of labor force and welfare participation	Comparison of change in individual AFDC-U case duration to change in matched counties	Random assignment at OJT recruitment	Comparison of forecast to actual annual caseloads	Pooled time series regression analysis of CWEP and non-CWEP county caseload trends	Caseload trend comparison
Comparison Group	Minimal WIN services	None	Pre ET-Choices caseload behavior; female family heads in other states	Non-CWEP counties	Applicants assigned to regular WIN services	Pre-CWEP caseload trends; also compare CWEP, non-CWEP counties	Non-CWEP counties	Washington state caseload
Maximum Duration of Observations	2 to 12 qtrs.	21 to 25 mos.	Not applicable	3 yrs., but results not reported	15 to 33 mos.	Not applicable	Not applicable	Not applicable
Benefit-Cost Analysis Perspective	Client, agency, taxpayer, social	None conducted	Budgetary costs, benefits	Agency	Client, agency, taxpayer, social	Taxpayer	Agency	None conducted
Comments	Rich program operated by noted CETA/JTPA agency	Process analysis of sequenced, rich program	Program emphatically voluntary, resource intensive; highly publicized	Time series evaluation failed	Similar to Maine, but fewer resources, smaller earnings impacts	Interesting and useful example of in-house evaluation	Time series analysis seriously flawed	Report of little value; example of inadequate in-house evaluation

TABLE 1

Summary Overview: The OBRA Demonstrations
(Part C: Pennsylvania to West Virginia)

	Pennsylvania CWEP	South Carolina CWEP	Utah EWEP	Virginia ESP	Washington AES	Washington CWEP	West Virginia AFDC-R CWEP	West Virginia AFDC-U CWEP
Extended Name			Emergency Work and Employment Program	Employment Services Program	Applicant Employment Services			
Location	Statewide	Beaufort and Spartanburg counties	Statewide; analysis data from 7 social service offices	Statewide; analysis data from 11 local welfare agencies	19 community service offices	Pierce, Spokane counties	Statewide	9 counties
Evaluator/Reference	Pennsylvania DPW (1986)	M.H. Clarkson Co.; South Carolina DSS (n.d.)	Janzen et al. (1987)	MDRC: Riccio et al. (1986)	Washington DSHS (Fred P. Fiedler) (1983)	Washington DSHS (Hal Nelson) (1984)	MDRC: Friedlander et al. (1986)	MDRC: Friedlander et al. (1986)
Eligible Household Types	AFDC-R,U	AFDC-R	AFDC-U replacement	AFDC-R	AFDC-R	AFDC-R	AFDC-R	AFDC-U
Target Clients	APPs, WIN eligible recipients	WIN eligible APPs, recipients from unassigned WIN pool w/o children <3	Same as AFDC-U	APPs, WIN eligible recipients w/o children <5 (<6 for study sample)	APPs w/o children <3	Mandatory WIN registrants selected by WIN office	APPs, NDMRs, CMRs	All WIN-mandatory AFDC-U adults
Program Components	AJS, JS, CWEP	Orientation, CWEP	JS, training, AB-GED, community work	AJS, JS-G, JS-I, CWEP, misc. training	AJS	CWEP, JS-G, JS-I	CWEP, WIN services	CWEP
Sequence	AJS/JS/CWEP	Orientation video/ CWEP	Not available	AJS/JS/options	AJS (project focused on early intercept)	Two tracks: CWEP; JS-G/JS-I	No sequence	No sequence
Requirement	Mandatory	Mandatory	Job of last resort	Mandatory	Mandatory	Mandatory	Mandatory	Mandatory
Operating Agency	SWA	CWD	SWA	CWD	SWA	SWA	SWA	SWA
Date of Program Initiation	4/82	10/82	1983	1/83	4/82	7/82	7/83	Early 1982
Date of Study Intake	10/83-4/84	10/82-3/84	Not applicable	8/83-9/84	6/82-1/83	Uncertain	7/83-4/84	3/83-4/84

(table continues)

TABLE 1
Part C (continued)

	Pennsylvania CWEP	South Carolina CWEP	Utah EWEP	Virginia ESP	Washington AES	Washington CWEP	West Virginia AFDC-R CWEP	West Virginia AFDC-U CWEP
Nature of Client Process Analysis	CI	CI	CI	CI	PITI	None	CI	CI
Nature of Comparison Group	Sample of unaffected recipients	Not applicable	1981 AFDC-U program	Impact controls	Comparable clients in control offices	None	Impact controls	Impact controls
Basis of Impact Analysis	None	None	Comparison to previous AFDC-U caseload	Random assignment at application or reevaluation	Local offices randomly assigned to experimental, control group	Random assignment of referrals to two experimental and one control group	Random assignment of clients to experimental, control groups	Counties randomly assigned to experimental, control group
Comparison Group	None	None	AFDC-U recipients	Controls receiving no services	Clients in control counties without AJS	Referred clients not served	Clients receiving WIN services, no CWEP	Clients in control counties without saturation funding
Maximum Duration of Observations	Not applicable	Not applicable	Not applicable	34 to 57 mos.	1 qtr. to 10 mos.	Not clear	22 to 29 mos.	22 to 35 mos.
Benefit-Cost Analysis Perspective	None conducted	Agency	Agency	Client, agency, taxpayer, social	Agency	Agency	Client, agency, taxpayer, social	Client, agency, taxpayer, social
Comments	Principally program description	Report of no value	Substantial programmatic innovation; unmitigated work requirement	Partial implementation failure; useful results nonetheless	Novel evaluation design; more analysis needed	Control, sample sizes inadequate; analysis flawed	Classic workfare in weak economy	Classic workfare in weak economy; unusual evaluation design

III. The Demonstrations: Process

While the requests published by agencies seeking to fund program evaluations often include calls for process analysis, the meaning and objective of process analysis is generally ambiguous. Here we concentrate on client-oriented process analysis that has two objectives: first, assessment of the consequences of a WEP for actual services received by clients, and second, the identification of the consequences of the program for the expected experience of persons who receive public assistance over a given interval of time. These two dimensions reflect our understanding of the avenues whereby WEP effects occur.

A Model of WEP Consequences

We begin with the traditional model of labor-leisure choice in the presence of an income support system. We consider the choices available to a single parent with children in a single time period. We initially assume that such a parent can obtain work at the minimum wage, but only at the cost of "leisure"--time not working--forgone.

This situation is illustrated in Figure 2, a standard labor-leisure diagram in which time not working during the period is measured on the abscissa and income is measured on the ordinate. Since A represents maximum leisure (i.e., no work), we can index utility levels by the height of the respective indifference curve at A: D is superior to C, which in turn is superior to B, and so on. If wages are fixed, the income available to the household from work alone is represented by an opportunities locus such as AM; the greater the wage, the steeper the slope of this line. Given opportunities AM, our household can do no better than to work h hours, gaining income K and utility C.

Now, suppose an income of level J is guaranteed by the welfare system. Even without work, the household will be better off now than before, since utility with welfare alone is on indifference

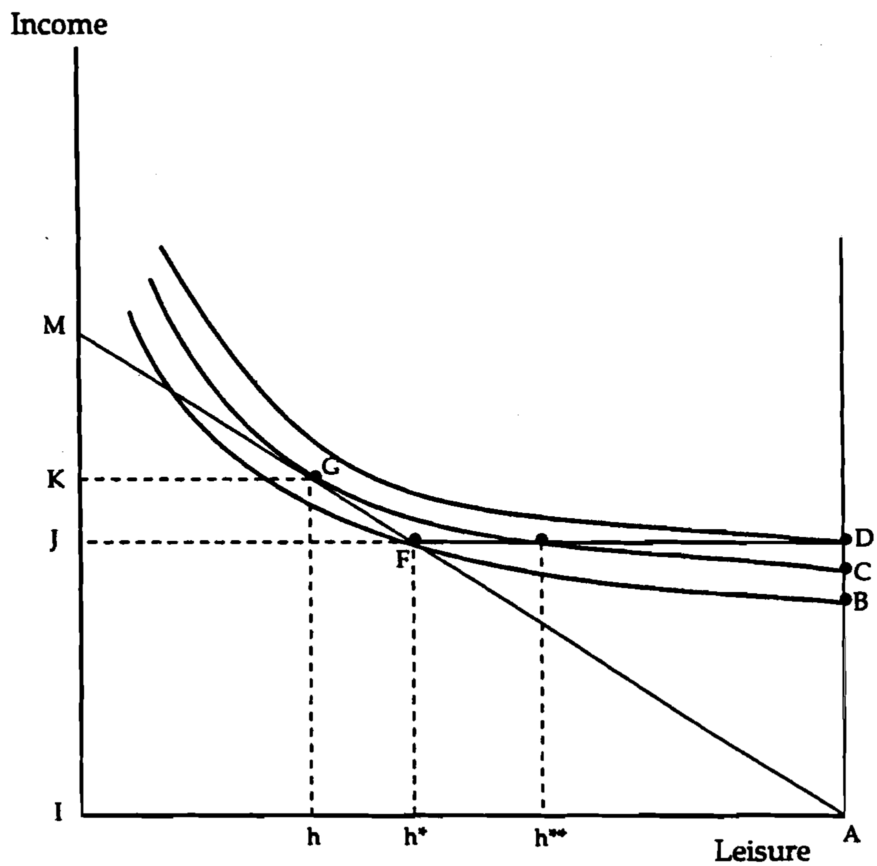

Figure 2: Client Behavior with Workfare

curve D, which is higher than C, the maximum utility attained in the absence of welfare. But what is the effect on labor supply? The effect of the introduction of the welfare system on labor supply depends on the treatment of earnings in welfare benefit calculations. For simplicity, assume that every dollar of earnings reduces the welfare payment by a dollar. In this case as leisure is reduced from A to h*--the "breakeven" labor supply level that produces income equal to the welfare guarantee--income does not change. At leisure levels below h* no welfare is received, but additional work does produce additional income.

With the welfare system, the new income opportunities locus is DFM. As drawn, the introduction of the welfare benefit will lead the household to drop out of the labor force altogether, for utility D (attained without working) is higher than C, the maximum level that can be achieved from work alone.

In simplest form, traditional work-for-welfare programs mandate that clients "work off" their benefits at the minimum wage. This is exactly what occurred, for example, in the West Virginia CWEP demonstrations. In the diagram, labor supply A-h* pays for the grant. Adding a work requirement equal to A-h* reduces the utility associated with welfare receipt from D to B. If welfare requires A-h* anyway, the fact that C exceeds B means that the household will prefer working outside of welfare over welfare receipt, even though becoming independent will mean working more (h is less leisure than h*). To the extent employment opportunities are available, the introduction of a work requirement will, in this model, increase labor supply and reduce the incidence of welfare receipt as long as work required exceeds some minimum amount A-h**. The higher the wage rate the recipient can obtain in the labor market, the lower this "tolerance threshold" will be.

The notion that welfare recipients have ready access to whatever level of employment they desire strains credulity. But more "realistic" specifications--making hours available the product of a random process or wages stochastic--do not change the conclusions. In either case, the introduction of a work requirement reduces the likelihood that any family at risk of dependence on welfare will actually take welfare up and raises the likelihood that those receiving assistance will leave because (obviously) such an obligation makes welfare receipt less desirable. How great this effect is depends upon the magnitude of the actual work requirement and its distastefulness, the incidence of employment opportunities, the potential wage rate, and the level of welfare benefits.

The WEPs summarized in Table 1 are not for the most part traditional "workfare" programs. But to varying degrees, Figure 2 still applies because such programs are time costly. The story of the diagram is that the time commitment raises the costs of welfare, thereby lowering its utility relative to the alternative(s). Moreover, from the standpoint of analyzing behavior, it is not the past but the future that counts. In Figure 2, our worker does not chose to work ex post, but ex ante--on the basis of the benefits and costs of welfare versus work. Thus, to analyze the consequence of a WEP for employment and welfare dependence, we need to measure the expected time costs.

Other WEP Effects

Other effects of WEPs complicate the picture just presented and dictate study of realizations, as well as expectations, in process analysis. We have labeled these the information, skills, and intertemporal incentive effects.

The typical WEP combines a time commitment with three broad classes of activities: (1) job search; (2) work, specifically CWEP, but also in other forms; and (3) a variety of general and specific skills training, ostensibly intended to increase wages available on the job. Better search might be expected to increase labor market information, raising the likelihood of finding a job offering enough hours or high enough wages to lead to independence from welfare. Skills training and perhaps even the experience gained from CWEP might be expected to raise the number of such jobs to which the client has access. The impact of the program derives from both the cost of the time required and the consequences of job search assistance and training for job information and access. But there is an important difference in the nature of time and search/training effects. Time effects may be realized ex ante, on the basis of expectations; search/training effects, on the other hand, can be realized only ex post. Only in The Music Man does expectation ("think the Minuet in G") lead to skills, and even in this case motivation took time and rehearsals. The consequence for process analysis is recognition that attention must be paid both to expectations--for time effect--and to realizations--for information/skill effects.

By definition, a single-period model such as that represented in Figure 2 misses any interrelationship between what occurs in one period and circumstances in the next. But clearly, the job search and skills components of OBRA demonstrations complicate analysis of client choices by introducing just such complexities. The time cost of welfare employment programs is expected to lower the likelihood of welfare dependence. But a WEP links pubic assistance with enhanced opportunities, both through better information and skills improvement. The promise of such opportunities may in the short run raise, rather than reduce, welfare dependence. However, if in the long run information is improved and skills training is efficacious, the likelihood of future employment and higher wages for a participant is increased, with reduced welfare dependence as a consequence.

While this model has been discussed in terms of a single client, welfare programs in fact provide income support to many, and the impact of work and training requirements are typically assessed in terms of consequences for all covered families. In terms of Figure 2, such a requirement would imply that covered families would be expected to work A-h* hours. But, in practice, the treatment is not a uniform requirement of A-h* hours, but rather greater risk of an A-h* requirement. This obligation becomes a risk rather than a feature because, for various reasons, some clients actually experience it and others do not. Whatever the selection process that creates such variation, its operation complicates replication of the treatment in other circumstances and vitiates attempts to draw inferences from particular WEPs about the consequences of similar programs elsewhere.

Selection effects can arise from both the program and client sides. Consider this example. Suppose that community work experience slots are available for only about one-quarter of all recipients in an experimental treatment group for whom outcomes are to be compared to a control group for which no CWEP is available. Suppose also that program operators have identified approximately one-quarter of all new recipients as malingerers and, through artful scheduling, have placed these persons in the available slots. If we define "malingerers" as people who fail to take advantage of known opportunities for self-support, we might expect both the immediate and longer-run consequences for dependence to be substantial if the suspected malingerers do, in fact, drop out of welfare.

Alternatively, suppose that one-quarter of all recipients are people seriously bent on finding employment who view community work experience as a vital opportunity to gain on-the-job experience and a track record suitable for citing in future employment interviews. Suppose, this time, that the community work experience jobs are given to volunteers. In this case, the consequence may be no immediate welfare reduction, but in the future employment effects may be observed as the motivation of these people, when combined with community work experience, pays off. Alternatively, it may be that the same group of go-getters would eventually have found employment anyway, but without CWEP it takes a little longer. In this case, the benefits of the CWEP-supported head start, as measured by experimental versus control group performance, attenuate over time.

Whatever happens, there is no reason to believe that the consequences of CWEP when targeted on malingerers will be the same as the consequence of providing employment only to volunteers. The implication is that selection is part of treatment. Since such selection effects are the result of design choices made by program administrators, prediction of the consequences of policy replication elsewhere requires predicting the behavior of administrators, as well as that of clients.

We emphasize the selection problem because we believe outside observers too often treat the welfare employment demonstrations as equivalent in structure to the negative income tax experiments of twenty years ago. However, the difference between the negative tax experiments such as the one conducted in Seattle and Denver and the OBRA experiments involves much more than just the distinction between treatments with cash and treatments with employment-oriented obligations. In the NIT experiments, <u>every</u> experimental received a cash payment when family income was sufficiently low. There was no selection. In the workfare experiments, the experience of treatment is far less certain and the opportunities for bias in inference are far more significant. Despite this difference in character and the possibility that subtle aspects of the workfare process may have profound consequences for site-to-site variation in the effects of programs of similar construction, most reporting continues to focus on impacts, and to treat process outcomes--or worse, process diagrams--as process inputs.

An Overview of Process Data

Our workfare model says that it is the expected value of the workfare time commitment and the realized incidence of search assistance, work experience, and training inputs that account for WEP effects. Therefore, a process analysis should measure these factors.

In principle, measurement of time commitment should be the most straightforward. There are two interrelated ways of looking at the time commitment. One is a point-in-time activity rate: on a typical day, what fraction of available work hours does a WEP client spend in a program-related activity? A problem with this measure is that activity levels may not be uniform over the various stages of the WEP process, and specified activity levels may be easier to achieve with some clients than others. As a consequence, the activity rate will be influenced by the composition of the client caseload as defined by both their personal characteristics and the point in the WEP process they have attained.

The superior alternative to assessing activity rate at a point in time is a cumulative measure. A cumulative participation rate would report the sum of time utilized by the program over a fixed interval of involvement as a proportion of all time available--say total work days. Normalization for variation in activity levels over the process sequence could be accomplished by measuring activity over a fixed interval commencing with, say, enrollment. The end result would be a number: a program that produced a cumulative activity rate of .5 for the first six months of client participation would involve, on average, half-time activity.

The point-in-time and cumulative participation rates described above are <u>outcomes</u>, and we have argued that it is <u>expectations</u> concerning time obligations that count. If an average cumulative participation rate masked considerable interclient variation in obligations imposed, then the average may not serve to identify well what an individual client familiar with the program expects. However, under most circumstances, it seems reasonable to believe that substantial differences in either participation measure across programs are closely associated with true differences in the time obligation imposed. If this is correct, an average measure of participation would seem the first objective of process analysis.

The second objective is measuring the incidence of component inputs. The first step in obtaining these measures is to identify the cumulative incidence of a component activity; the second is to evaluate the relation of variations in utilization of this input to client characteristics. Ideally, this would allow one to say something like the following: "A client beginning the Yonoheeta County Jobs-Are-Good program can expect to devote about 50 percent of her time on public assistance during the subsequent six months to JAG activities; if she has recent labor market experience and no specific employment barriers, she will receive job search assistance and 80 hours of training in business mathematics." Note that this is a conditional statement: it identifies the services a client will receive if she stays in the program. If when presented with this prospect, she chooses to drop out, that is a behavioral response. We cannot begin to understand this response without identifying the stimulus. This is one of the things process analysis is about.

Study of both participation and component utilization in OBRA-related programs is complicated by a number of factors. For most of the studies we have surveyed, the major complication arises because often the process evaluations are embedded in the impact evaluations. As was discussed in Section II, many of the impact evaluations are based on samples that are drawn not at the point clients begin WEP activities, but at the point they apply for welfare. As a result, a significant proportion never receive public assistance at all and, indeed, may never be given the opportunity to participate in anything WEP-related. Participation incidence calculated from such circumstances reflects both the level of activity achieved within the program and the amount of fallout between the point of application and WEP initiation. While fallout may in part be a response to the WEP participation obligation, this fallout (and consequent nonparticipation) should not be accounted as evidence of an agency failure.

A second complicating factor is turnover: whatever the interval chosen for assessing cumulative participation, a significant number of even those clients who make it into the process will leave welfare before this level of exposure is achieved. As a result, unadjusted measures of the incidence of receipt of particular services do not provide good measures of program process. Normalization for actual time "at risk" is essential.

Sanctioning also complicates interpretation of participation data. Although the incidence of sanctioning is often taken as a measure of the stringency of the obligation imposed by welfare employment, it is an <u>outcome</u>, not an <u>input</u>. It is the result of the interaction of client behavior and procedures established by the sponsoring agencies. The input of interest is agency policy and behavior. If the WEP agency pays careful attention to client compliance with system procedures and moves quickly in response to evidence that clients are evading program obligations, then actual sanctions applied may go up. Alternatively, they may go down as clients respond to agency behavior by adjusting their own. Moreover, there is no necessary connection between the level of sanctioning and the activity rates a program achieves. For example, a well-run program may, with appropriate attention to motivation and substance, actually lower client recalcitrance while raising activity rates. Aside from this, sanctioning may lower participation rates. This is because the hearings-and-appeal process guaranteed adult welfare recipients who do not comply with WEP regulations takes time, and while it is going on, activity may cease. This lowers participation, however measured. Ideally, participation measures should be adjusted to eliminate this effect, but they never are.

Our view is that rather than focus on sanctioning alone, process analysis should attempt to identify what the WEP agency considers to be the client's obligations and the agency's operating procedure for dealing with noncompliance. Given this information, more may be learned from the outcomes--the extent of failure to comply, the frequency with which sanctioning procedures are invoked, and the rate at which sanctions are actually applied.

Sample Process Information

We found no process evaluations which report either point-in-time or cumulative participation rates in the form we suggest is appropriate to the theory of WEP effects. However, many reports included pieces of a process analysis, and these fragments provide useful information for interpretation of impacts. In Tables 2 through 5, we provide several examples of these data. Our approach in the discussion that follows is first to describe these tables and then to identify what appear to us to be the important conclusions to be drawn.

Tables 2 through 4 provide illustrative examples of the process data that are available. We have divided these data into three groups defined on the basis of the type of household (AFDC-R or U) and the status of the client at the point of selection for use in evaluation. Table 2 covers data for clients from AFDC-R households selected at the point of application for welfare. Table 3 reports information for newly determined and continuing mandatory recipients, also from AFDC-R households. The last table, Table 4, presents data for AFDC-U applicant samples. For a specified time interval, each table provides the cumulative incidence of: (1) welfare participation, (2) component participation, (3) deregistration, and (4) sanctioning. Because of data limitations, the sample restrictions are only loosely applied: for example, the information on the California GAIN program in Table 3 includes both applicants and continuing recipients. All but the results for the Washington Applicant Employment Services and the Pennsylvania CWEP demonstrations are for programs evaluated by MDRC.

Interpretation of these tables and the shortcomings of the available data are usefully illustrated by examining one of the columns in detail: the first column in Table 2, which covers the first nine months following assignment to control or experimental status for AFDC-R applicants in the Arkansas WORK program. Over this period, only 59 percent of the applicants actually received any welfare payments and 38 percent participated in some WEP component, with only 4 percent receiving a work experience assignment. Since registration in WORK is a precondition of welfare eligibility in this system, it was possible for more people to register for WORK than were ever certified as eligible for welfare. All those applicants who registered but never achieved welfare eligibility, were eventually deregistered, and they, as well as applicants who did receive welfare but for fewer than nine months, are included in the 66.5 percent reported deregistration rate. The remaining columns are interpreted similarly, but with the adjustments indicated by the notes.

In summarizing available information on client process, we would have liked to stay as closely as possible to our model of what should be collected, especially with respect to the cumulative and point-in-time participation measures. However, as the Arkansas data indicate, this was rarely possible. Instead, for incidence of participation we generally had only a measure of the proportion of a client group who received welfare or participated in one or more services over some fixed period of time. This is what we call a "cumulative incidence" (CI) measure in Table 1. CI is not a measure of the share of the interval devoted to specific or general WEP activities.

Moreover, these reports do not adjust cumulative participation incidence for time the client was actually at risk of program participation. Thus in the Arkansas case, we know that 38 percent of applicants participated in a program activity, but we do not know the conditional likelihood of

TABLE 2

Cumulative Incidence of Participation Data for AFDC-R Applicants in Selected OBRA Demonstrations

	Arkansas WORK	California San Diego EPP/EWEP	California San Diego SWIM	Maryland Baltimore Options	Maryland BET	Virginia ESP	Washington AES	West Virginia
Household type	AFDC-R	AFDC-R	AFDC-R	AFDC-R	AFDC-R	AFDC-R	AFDC-R	AFDC-R
Status on assignment	Applicants assigned at WIN registration	Applicants assigned at AFDC application	Applicants assigned at SWIM registration	Applicants assigned at AFDC application	Applicants assigned at AFDC application	Applicants assigned at AFDC application	Applicants assigned at AFDC application	APPs, NDMRs assigned at WIN registration
Follow-up time period (months)	9	9	12	12	12	9	1	9
Received welfare	58.9	81.2 (first 6 mos.)	85.0	n.a.	n.a.	69.3	n.a.	90.8
Ever participated	38.0	47.3	56.7	53.0	45.3	41.7	n.a.	16.7
JS-I	25.5	n.a.	n.a.	n.a.	n.a.	29.9	58.6	
JS-G	23.5	45.0	n.a.	n.a.	n.a.	9.8	14.3	
JS (any)	n.a.	n.a.	45.7	32.7	43.0	37.2	n.a.	
CWEP/WWE	4.0	13.0	14.8	14.6	14.8	3.2		16.7
OJT			0.8					
Training	n.a.	n.a.	18.1	17.7	11.7	8.8		
Deregistered	66.5	59.7	70.0	57.2	60.2	54.8		69.5
Due to sanction	n.a.	7.5	12.8	n.a.	n.a.	8.8		1.4
Control experience	Virtually no services	Virtually no services	Virtually no WIN services; see text	Available data not comparable; controls received very few services	No controls	Comparable data not available; no ESP services	Comparable data not available; apparently few services	Comparable, except no CWEP

(table continues)

TABLE 2 (continued)

	Arkansas WORK	California San Diego EPP/EWEP	California San Diego SWIM	Maryland Baltimore Options	Maryland BET	Virginia ESP	Washington AES	West Virginia
Notes		CWEP data are for JS/EWEP track only. Other data are two-track averages	Welfare receipt refers to qtrs. 2 through 5			Welfare receipt refers to qtrs. 1 through 4		Welfare receipt refers to qtrs. 2 through 7
Source	Friedlander et al. (1985) pp. 75, 95	Goldman et al. (1986) p. 194	Hamilton (1988) pp. 109-10, 204	Friedlander et al. (1985) p. 71	Friedlander et al. (1985) p. 196	Riccio et al. (1986) pp. 56, 98	Washington DSHS (1983) p. 8	Friedlander et al (1986) pp. 73, 216

Notes: Unless otherwise indicated, all numbers are percentages of original sample which had, by the end of the indicated follow-up period, participated in the indicated activity.

n.a. = not available

Cells are left blank where statistic is inapplicable to demonstration.

TABLE 3

Cumulative Incidence of Participation Data for AFDC-R Recipients in Selected OBRA Demonstrations

	California San Diego SWIM	California GAIN	Illinois WDP	Maryland Baltimore Options	Maryland BET	Pennsylvania CWEP	Virginia ESP	West Virginia
Household type	AFDC-R	AFDC-R	AFDC-R	AFDC-R	AFDC-R	AFDC-R	AFDC-R	AFDC-R
Status on assignment	Recipients assigned at SWIM registration	Applicants, recipients assigned at GAIN registration	Applicants, recipients assigned after AFDC established	Recipients assigned at WIN referral	Recipients assigned at WIN referral	Applicants, NDMRs	Recipients assigned at WIN referral, review	CMRs assigned at reappraisal
Follow-up time period (months)	12	6	9	12	12	6	9	9
Received welfare	94.9	n.a.	99.8	n.a.	n.a.	n.a.	97.2	99.2
Ever participated	69.6	26.6	47.4	52.7	67	n.a.	69.4	
JS-I	n.a.	n.a.	35.9	n.a.	n.a.	n.a.	47.4	
JS-G	n.a.	n.a.	0.0	n.a.	n.a.	n.a.	18.0	
JS (any)	53.8	10.2	35.9	25.4	55.3	n.a.	60.2	
CWEP/WWE	22.0	0.7	7.3	26.9	27.7	4.8	13.6	27.0
OJT	0.6					n.a.		
Training	28.4	26.3	17.4	22.9	21.3	n.a.	13.5	
Deregistered	55.8	28.9	55.6	28.8	40.4	n.a.	34.0	30.6
Due to sanction	9.1	0.5	11.8	n.a.	n.a.	n.a.	17.7	2.0
Control experience	Virtually no WIN services; see text	Control not yet established	Training incidence comparable to treatment group	Available data not comparable; controls received very few services	No controls	No controls	Comparable data not available; no ESP services	Comparable, except no CWEP

(table continues)

TABLE 3 (continued)

	California San Diego SWIM	California GAIN	Illinois WDP	Maryland Baltimore Options	Maryland BET	Pennsylvania CWEP	Virginia ESP	West Virginia
Notes							Welfare receipt refers to qtrs. 1 through 4	Welfare receipt refers to qtrs. 2 through 7
Source	Hamilton (1988)	Riccio et al. (1989) p. 85	Friedlander et al. (1987) pp. 45, 81	Friedlander et al. (1985) p. 71	Friedlander et al. (1985) p. 196	Pennsylvania DPW (1986) p. 19	Riccio et al. (1986) pp. 56, 98	Friedlander et al. (1986) pp. 73, 218

Notes: Unless otherwise indicated, all numbers are percentages of original sample which had, by the end of the indicated follow-up period, participated in the indicated activity.

n.a. = not available

Cells are left blank where statistic is inapplicable to demonstration.

TABLE 4

Cumulative Incidence of Participation Data for AFDC-U Samples in Selected OBRA Demonstrations

	California San Diego EPP/EWEP	California San Diego SWIM	Maryland Baltimore Options	Maryland BET	Pennsylvania CWEP	West Virginia
Household type	AFDC-U	AFDC-U	AFDC-U	AFDC-U	AFDC-U	AFDC-U
Status on assignment	Applicants assigned at AFDC application	Applicants assigned at SWIM registration	Applicants assigned at AFDC application	Applicants assigned at AFDC application	Applicants assigned at AFDC application	APPs, NDMRs assigned at WIN registration
Follow-up time period (months)	9	12	12	12	6	9
Received welfare	77.8 (first 6 mos.)	86.2	n.a.	n.a.	n.a.	89.5
Ever participated	51.6	62.6	50.4	37.9	n.a.	
JS-I	n.a.	n.a.	n.a.	n.a.	n.a.	
JS-G	50.4	n.a.	n.a.	n.a.	n.a.	
JS (any)	n.a.	56.7	34.8	37.9	n.a.	
CWEP/WWE	16.7	16.9	15.7	15.5	14.4	52.9
OJT		0.7			n.a.	
Training	n.a.	12.7	11.3	0.0	n.a.	
Deregistered	67.0	72.6	76.5	70.7	n.a.	79.4
Due to sanction	5.6	9.3	n.a.	n.a.	n.a.	5.5
Control experience	Virtually no services	Virtually no WIN services; see text	Available data not comparable; controls received very few services	No controls	No controls	n.a.

(table continues)

TABLE 4 (continued)

	California San Diego EPP/EWEP	California San Diego SWIM	Maryland Baltimore Options	Maryland BET	Pennsylvania CWEP	West Virginia
Notes	CWEP data are for JS/EWEP track only. Other data are two-track averages	Welfare receipt refers to qtrs. 2-5; data include APPs, NDMRs, CMRs recipients				Welfare receipt refers to qtrs. 2-7; data include APPs, NDMRs, CMRs
Source	Goldman et al. (1986) p. 194	Hamilton (1988) pp. 109, 110, 204	Friedlander et al. (1985) p. 71	Friedlander et al. (1985) p. 196	Pennsylvania DPW (1986) p. 19	Friedlander et al. (1986) pp. 165, 190

Notes: Unless otherwise indicated, all numbers are percentages of original sample which had, by the end of the indicated follow-up period, participated in the indicated activity.

n.a. = not available

Cells are left blank where statistic is inapplicable to demonstration.

participation, that is, the likelihood that any client found eligible for public assistance and receiving benefits for, say, two months, would be involved in job search. Yet another shortcoming results from counting clients as having participated in a program during a month if he or she attended on a single day. For example, one day of job search during the first nine months after AFDC application could get a client counted in the 38 percent of recipients assigned to WORK in Arkansas who "ever participated." The table would not change at all had that same client worked intensively for eight weeks at searching for potential employers. As is probably obvious, no adjustment for time lost to the sanctioning process is possible.

The remaining data in Table 2 are interpreted as those for the Arkansas WORK program were, with the adjustments indicated by the notes. The Washington AES experiment is unusual for its short time-horizon: one month. This experiment dealt only with services provided welfare applicants before they received their first checks and provided only mandatory job search assistance.

Table 3 covers results for programs enrolling AFDC-R recipients. Here we obtain a somewhat clearer picture of the degree of turnover within the WEP client population than was available from the applicant data, since almost all participants were already receiving public assistance payments when assignment occurred. While it is difficult to generalize from data spawned by such a diverse set of programs as those represented in Table 3, it appears that, in all sites, at least one-fourth of AFDC-R recipient participants had deregistered from WEP within the follow-up period. Unfortunately, we cannot be confident about what "deregistration" means, since it can come about both as a result of termination of welfare and as a result of changes in circumstances--pregnancy, for example--that remove a person from WEP eligibility.

For reports in which process data are available separately for applicants and recipients, it appears that participation incidence is higher for recipients than for applicants. This probably reflects differences in the greater participation opportunities had by recipients--who at the time of WEP registration were already within the welfare system--compared to applicants--who often did not attain welfare eligibility at all.

One exception to this generalization is the Baltimore Options program; here the share of applicants and clients who ever participate in any service is virtually the same--53 percent. But it does appear that the Options program treated the two client types differently. Twenty-seven percent of NDMR clients participated in work experience, while only fifteen percent of the applicant group did. However, interpretation of this difference must be tentative because we cannot control for time at risk. All other things equal, it may well be that clients from the applicant pool are more, rather than less, likely than clients from the NDMR pool to be assigned to CWEP during the time they remain in WEP. But deregistration rates are twice as high among applicants as they are among the recipients in the Options program. Therefore, the greater incidence of CWEP participation among the NDMRs may merely reflect the fact that these women were more likely to receive and remain on welfare.

Table 4 reports data for selected AFDC-U client samples that are comparable to those reported in Table 2. The smaller number of OBRA demonstrations providing data on this group makes generalization very risky. Here, as is true of most other AFDC data, it appears that deregistration occurs more rapidly for U-clients than for R-clients.

For evaluation of impact results, as well as to assess the degree to which a WEP changes the orientation of a welfare program, it is important to identify the experience of the control group. A summary judgment of the nature of the control process appears in Tables 2 through 4 for each of the demonstrations summarized there. Table 5 provides specific information on the services controls received for the four OBRA demonstrations for which the necessary data were available.

Selection for inclusion in Tables 2 through 5 was done solely on the basis of information availability. For many of the demonstrations, such information is not presented, and we must rely on

TABLE 5

Service Receipt by Controls, for Selected OBRA Demonstrations

	California San Diego EPP/EWEP		Illinois WDP		New Jersey Grant Diversion		West Virginia	
Household type	AFDC-R		AFDC-R		AFDC-R		AFDC-R	
Status on assignment	Applicants assigned at AFDC application		Applicants, recipients assigned after AFDC eligibility established		Recipients assigned after screening		Applicants, NDMRs assigned at WIN registration	
Group	Experimental	Control	Experimental	Control	Experimental	Control	Experimental	Control
Received welfare	81.2 (first 6 mos.)	82.7 (first 6 mos.)	99.8	99.8	97.6	97.2	90.8	88.7
Ever participated	47.3	6.1	47.4	16.9	82.9	70.5		
JS-I	n.a.	n.a.	35.9	2.2	61.5	46.5		
JS-G	45.0	0.8	0.0	0.0	18.6	22.5		
JS (any)	n.a.	n.a.	35.9	2.2	61.5	66.9		
CWEP/WWE	13.0	0.0	7.3	0.0	6.4	8.5	16.7	1.0
OJT					42.3	0.0		
Training	n.a.	n.a.	17.4	15.9	18.2	17.1		
Deregistered	59.7	52.0	55.6	46.5	42.6	39.8	69.5	68.6
Due to sanction	7.5	0.7	11.8	7.6	5.7	6	1.4	0.8
Follow-up time period (months)	9	9	9	9	12	12	9	9
Notes	CWEP data are for job search/work experience track only. Other data are two-track averages		CWEP data are for job search/work experience track only. Other data are two-track averages		Training refers to JTPA		Welfare receipt refers to qtrs. 2-7	
Source	Goldman et al. (1986) p. 194		Friedlander et al. (1987) pp. 45, 81		Freedman et al. (1988) p. 150		Friedlander et al. (1986) pp. 73, 216	

Notes: Unless otherwise indicated, all numbers are percentages of original sample which had, by the end of the indicated follow-up period, participated in the indicated activity.

n.a. = not available

Cells are left blank where statistic is inapplicable to demonstration.

general descriptions of procedures to make inferences about the incidence and timing of WEP activities. Nonetheless, our reading suggests that five important conclusions emerge from process analyses of the nature of OBRA demonstration inputs and their consequences for client experiences:

(1) Job search is the major treatment. Of the OBRA demonstrations and client subsamples included in Tables 2 through 4, only about half of the treatment group members participated in any component activity over an interval of nine to twelve months following assignment. For those who participated in anything, job search was the most common by far. With the exception of West Virginia's CWEP saturation experiment, the odds are generally less than one-in-five that a member of the treatment group in one of these experiments will actually become involved in CWEP or training during the nine to twelve months following WEP registration.

(2) Turnover is substantial. Table 2 indicates that for the demonstrations for which data were available, over half of applicants eventually deregistered from the WEP. Interpretation of this number is complicated, however, by the fact that so many never apparently achieved welfare eligibility. Deregistration rates are reasonably high for AFDC-R recipients, too, as Table 3 suggests. Recall, however, that deregistration can come about for reasons other than employment; our understanding of OBRA outcomes would be aided by better information on the reasons for these transitions.

(3) Sanctioning occurs. In part because of the celebration by Massachusetts officials of the voluntary character of the ET-Choices program, considerable debate has developed over the past few years on the utility of participation enforcement. There are in the demonstration reports frequent references to difficulties encountered in securing client cooperation and attendance in WEP-related activities (as well as to the eagerness of some clients to receive WEP services). In practice, a small, but significant, proportion of participants in these

programs have been sanctioned. Very low sanctioning rates for California's GAIN program seem to reflect both start-up problems and difficulties encountered in administration of the program's cumbersome sanctioning procedures.

(4) Service receipt is often selection dependent. A very low incidence of receipt of particular services in complex programs--for example, training in the Baltimore Options and Virginia ESP demonstrations--means that someone's discretion is exercised in determining who gets what. As a result, whatever impacts are observed are potentially the product of both the services delivered and an unobserved assignment procedure. This problem is least significant for the minimum service programs that appear to have achieved high degrees of participation--for example the job search segments of the original San Diego program, the Illinois WDP, and the San Diego SWIM demonstration.

(5) Circumstances of controls vary. In Table 5, it is obvious that for the San Diego, Illinois, and West Virginia demonstrations control services were minimal. Indeed, MDRC's representation of the results of the initial San Diego experiment as an indication that the EPP/EWEP program was superior to WIN was challenged by the California Legislative Analyst's Office on the grounds that, as inadequate as services under WIN were, they were never as bad as what the EPP/EWEP controls received (Warren, 1985). However, in other demonstrations, controls receive more. This is particularly true in component-oriented demonstrations such as Maine's TOPS and the New Jersey Grant Diversion program. Controls in the New Jersey experiment received significant services; indeed, data presented in Table 5 indicate that over the twelve-month interval following assignment, the incidence of job search services among controls was larger than the incidence among experimentals. In the Virginia ESP experiment, controls received sufficient training to contaminate one part of the outcome, leading to its exclusion from analysis.

The problem of assessment of control services has increased with the greater targeting of services provided through the Job Training Partnership Act (JTPA) network. By regulation, all AFDC participants are targeted by JTPA providers. Evidence from the San Diego SWIM demonstration, the California GAIN evaluation, and the Minority Single Parent demonstration (a random assignment intervention evaluation funded in six areas by the Rockefeller Foundation) indicates that unserved controls do often find training opportunities elsewhere (Mathematica Policy Research, 1989). The upshot of these discoveries is that any evaluation of the net impacts of welfare-related services must develop methods for determining what services are received from non-welfare-related organizations by members of both control and experimental groups. MDRC has been quite diligent in its efforts to identify such service use in the GAIN and SWIM demonstrations, with useful results.

The Incidence of Participation

At the beginning of this section, we emphasized the multiple avenues of possible effects of WEP on recipient experience. By and large, the focus of the kinds of cumulative incidence data typically collected by MDRC is upon the search and human capital effects--that is, upon services actually received. Far less attention has been paid to measurement of the time cost of WEPs or to identifying the expected sequence of services and obligations a client entering a WEP (for example, the multiple-service Baltimore Options program) and remaining eligible might be expected to receive and the likely timing of this sequence. For analyzing the consequences for behavior, it may also be important to study what clients <u>believe</u> will happen in welfare, especially in circumstances in which the actual experience of program participants suggests something quite different. What do clients see in their future when they cross the WEP threshold?

For many of the OBRA demonstrations, it is not clear what participants were led to believe about the program process. In the first of the San Diego demonstrations, for example, clients in the job search/CWEP track were not told that CWEP was in their future until they were well into job search. Other demonstration reports provide little information at all about what recipients were told. In the absence of better information, the best approach would be to study the actual experience of clients who remain program eligible. What is the expected sequence of activity and time commitment generated for a recipient, given continued participation? For the most part, we do not know. Of course, because of the now-familiar problem of selection bias, answering this question is not simple. Those recipients who stay WEP eligible are themselves exceptional. Fortunately, statistical techniques exist for addressing selection bias. Unfortunately, they have yet to be applied to study of WEP process.

A somewhat cruder, but nonetheless useful, approach to investigation of WEP participation is to focus on the likelihood of activity at a point in time, as opposed to the cumulative incidence of activity as reported in Tables 2 through 4, or the measure of participation rate we described earlier. Obviously, for example, if all eligibles in a WEP do something every month, then the prospect for new participants who stay on welfare is most likely the same. These point-in-time incidence (PITI) measures are available for a small subset of the OBRA demonstrations. For the most part, they are calibrated on a monthly basis, so any activity within a month makes that month count as active. Data in the MDRC report on Arizona's WID program, for example, indicate that over the interval covered, on average, about 25 percent of WID eligibles were active in each month. This number is comparable to the 23 percent achieved in the Minnesota CWEP program. Higher rates were achieved in both the San Diego SWIM program and the West Virginia CWEP demonstrations, but in both of these cases saturation was one objective of the effort. Data reported by Hamilton (1989) indicate that, during a twelve-month interval, both AFDC-R and AFDC-U clients were active in SWIM during about half of the months for which they were registered and eligible for activity. West Virginia achieved a 59 percent activity rate using a very simple CWEP program targeted at men in AFDC-U.

Typically, a comparable activity number is not available for controls, but West Virginia is again an exception: measured on the same basis, rates of CWEP participation among controls were reported to be 34 percent.

The welfare employment program established by the Family Support Act of 1988, the Job Opportunities and Basic Skills Training program (JOBS), includes a participation goal. By 1995 states are to achieve a monthly activity rate of 20 percent. Compared to the participation rates reported for the OBRA demonstrations, the requirement defined in program regulations appears at once generous and restrictive. It is generous in that it is low, phase-in is gradual, and the statutes allow the numerator of the ratio to include both mandatory participants and volunteers, while the denominator covers only those required to participate. It is restrictive in that only recipients who are scheduled for an average of twenty hours per week in approved activities over the month and who make satisfactory progress while actually attending at least 75 percent of their scheduled hours are counted in the numerator. In this respect the JOBS standard is comparable to the participation rate measure we have suggested. See Wiseman (1991b) for details on the limits to this comparison. In contrast, the OBRA reports generally count any activity, no matter how brief, as participation. MDRC's evaluation of the San Diego SWIM program provides the most comprehensive analysis of participation; there, a client was counted as a participant if he or she was active in some program component during only one day a month. The CI measures reported earlier in this section count participation as having occurred if a client was active during only one day over an interval as long as twelve months. Thus, the OBRA demonstrations give only weak evidence on the feasibility of mandating rates of participation for agencies. It is clear from the OBRA reports, however, that, in general, operation of WEPs poses significant organizational and management problems. One of us has argued that such programs in fact provide more of a "work test" for bureaucracy than for the clients (Wiseman, 1987).

IV. The Demonstrations: Impact

Given process, did the OBRA demonstrations achieve the anticipated effects? As indicated by Table 1, impact analyses, which were designed to address such a question, were conducted as part of all but a few of the OBRA demonstration evaluations. In this section, we examine these impact analyses in some detail.

We begin this section by briefly enumerating the demonstration outcomes or impacts that the impact analyses attempted to measure. We then describe the numerous ways in which the evaluation designs associated with the various demonstrations influenced the estimates of these impacts. This is followed by a discussion of the statistical techniques used to measure the direction and magnitudes of the impacts and some methodological problems associated with the use of these techniques. Finally, findings from the impact analyses in which we place some confidence are reported in a fourth subsection.

Much of the discussion in this section is concerned with methodological issues that arose in conducting the impact analyses found in the evaluations we reviewed. Such a discussion inevitably focuses on evaluation problems and shortcomings. In doing this, we do not mean to imply that we consider the OBRA impact analyses to be of low quality. Some are, as we indicate later. But all those conducted by MDRC, as well as several of the others, are very well done. Because of data limitations and resource constraints, however, all the impact analyses suffer from certain limitations. Moreover, trade-offs occurred and choices had to be made. We attempt to clarify these limitations and trade-offs and make explicit some methodological issues that receive little discussion in the evaluation reports themselves, so that findings from the OBRA impact analyses can be better understood.

Impacts of Interest

The goal of most of the OBRA demonstrations was to improve the labor market performance of AFDC recipients and, thereby, to reduce their dependency on welfare. The measures of impacts used in the evaluations reflect this goal. These measures include changes in the employment status, earnings, welfare status, and transfer payments of AFDC recipients, and, at a more aggregate level, changes in caseload size and welfare expenditures.

In measuring and reporting on these impacts, evaluators of the OBRA demonstrations have placed the greatest emphasis on effects on the welfare population as a whole, or at least that part of the population eligible for treatment under the demonstration programs. However, because resources to fund programs aimed at reducing welfare dependency are scarce, some effort has been made, especially in those demonstrations evaluated by MDRC, to determine whether treatment impacts differ among various subgroups. If they do, then program resources can be targeted on members of those subgroups for which the treatment has the greatest payoff.

Researchers have been particularly interested in the relationship between program impacts and the characteristics that the growing body of research on welfare turnover has associated with duration of dependence. Simplified somewhat, this work suggests that welfare cases differ with respect to the probability that they will close over any given interval following opening. Since some cases ("movers") leave more rapidly than others ("stayers"), there tends to be a systematic difference between the average characteristics of all cases that ever open over, say, a year, and the characteristics of those cases open on any given date. Specifically, the point-in-time caseload tends to have a higher proportion of "stayers" than is true for any cohort of new openings. Compared to movers, stayer cases account for a disproportionate share of welfare costs. This implies that if the objective of welfare employment policy is to reduce welfare costs, then it would be particularly useful to find some treatment that transformed stayers into movers.

Termination rates for welfare cases appear to be related to the characteristics of the adults present, the size of the family and the ages of children present, the nature of the public assistance system, and the economic environment. Part of this can be summarized in terms of whether or not adults are "job ready," that is, they have the skills and experience that give them employment opportunities. But the available research suggests also that many of the influences on termination probability are not readily quantified and are certainly not reported in case files or covered in questionnaires. In such cases, it is "time that tells" who is a mover. These considerations focus attention on differences between employment program effects upon the job ready and those who are not job ready, and also upon differences in program effects upon new entrants--presumably more likely to include movers--and upon long-term recipients, a group dominated by stayers. Study of effects for new entrants is common also for administrative reasons; it is somewhat easier administratively to initiate a program by concentrating on new applicants as opposed to the entire population of eligible recipients.

An additional important subgroup comparison is between AFDC recipients with younger children and those with only older children. To some extent this, too, is a distinction between movers and stayers, since women with preschool children might be expected to face greater barriers to employment and to view the opportunity costs of working outside the home as more substantial. Historically, WIN was only mandatory for single parents who had no children under six years of age. Single parents with children between three and six could, however, be required to participate in the OBRA demonstration programs, and as Table 1 indicates, such a requirement was, in fact, made part of several of the demonstrations. It has for some time been clear that rising rates of labor force participation among women with preschool children would lead to political pressure for extending work requirements to welfare recipients among this group. Consequently, evaluation of program impacts for this subgroup is of considerable interest. Finally, potentially useful comparisons can be made of impacts on AFDC recipients in different geographic areas--for example, urban and rural areas or areas with high and low levels of unemployment.

In interpreting impact estimates from the evaluations, it is important to keep in mind the population to which they are applicable. This point has received virtually no emphasis in the OBRA evaluation literature, but it has an important bearing on judgments concerning how "large" or "small" the measured impacts are. The reason for this is that impacts are inevitably reported as a change in a ratio--for example, a dollar increase in earnings per person or a percentage point reduction in the number of persons receiving welfare. The demonstration programs, however, were usually targeted at only a subset of the full AFDC population--often those who were WIN-mandatory. And of the targeted population for a particular program, typically only a subset actually participated. Hence, there were three obvious alternative candidates for use in the denominators of the impact ratios: all recipients, all members of the target population, and only those persons who actually participated in the demonstration program. Since each of these groups is successively smaller than the preceding one, and at least the human capital and information effects are concentrated among participants, measured impacts will become increasingly large as one moves from the first candidate to the second and then to the third.

These three alternative impact denominators can be used to address three quite distinct policy questions:

(1) Can the demonstration treatments substantially reduce the size of the welfare caseload or total expenditures on welfare and can they appreciably increase earnings among the welfare population?

(2) Did the demonstrations have substantial impacts on those at whom the treatments were directed?

(3) Did the demonstrations have substantial impacts on those who actually received the program treatments?

Each of the OBRA evaluations tended to focus exclusively on only one of the three questions. Several--particularly those that emphasized program effects on overall caseload size--measured impacts relative to the entire recipient population. In general, the demonstrations upon which MDRC focused were designed to answer question 2. As will be seen, this choice was made, in part, for solid statistical reasons. Often the demonstrations are interpreted as if they were designed to answer the third question. It is this one that is of particular interest if participation in the demonstration services is itself a policy variable. However, only a few evaluations attempted to measure impacts for only those who actually received treatment services under the programs, and as discussed later, these studies were highly flawed.

We address these issues in more detail below. Our point here is that in examining findings from the OBRA evaluations, it is important to recognize that the effect of the choice of denominator on the magnitudes of the impact estimates is potentially quite substantial. In the case of the Virginia Employment Services Program, for example, 46 percent of the AFDC caseload in the eleven participating counties was excluded from the demonstration's target group (MDRC: Riccio et al., 1986, p. 29), and of the target group, 42 percent never participated in any of the treatment components.

Evaluation Design Issues

Experimental Versus Nonexperimental Designs

As indicated in Table 1, several alternative evaluation designs were used in attempting to assess the various OBRA demonstrations. The most prevalent of these was an experimental design in which the targeted adults were randomly assigned to one or the other of two types of groups: treatment groups and comparison groups. In principle, only persons in the treatment group were to participate in the activities and receive the services being tested in the demonstration. (In practice, there were some exceptions to this rule, but these exceptions were usually minor.) Because of random assignment, the characteristics of the cases in the two types of groups should have been similar and differences that occurred by chance could be at least partially adjusted for through standard statistical techniques. Consequently, the impacts of the treatments could be measured by simply comparing members of the two types of groups in terms of outcomes of interest.

This experimental design, which was used in all but one of the MDRC impact analyses and in several of the other evaluations as well, can be usefully contrasted to some of its alternatives. The most important of these alternatives, which we shall call the "treatment-comparison site" design, has been utilized in evaluations where the demonstration sites covered only part of a state. Under these circumstances, a treatment group can be constructed of cases located in the demonstration sites and a comparison group of cases located in other sites. (Similarly, in the case of a statewide program, the state itself can be treated as a "site" and compared to other states.) The obvious shortcoming of this evaluation design is that the treatment and comparison sites may differ from one another in ways other than treatment status, including local labor market conditions, characteristics of the AFDC caseload, and approaches to and philosophy in administering welfare. Differences in outcomes that are observed in comparing cases at the two types of sites may be attributable to these other factors, rather than to the demonstration treatment.

This shortcoming can be at least partially overcome, however, by using statistical techniques to adjust for those site differences that can be observed and measured. This was not done in all the studies we reviewed that used the treatment-comparison site design, however, and even when it was, there was no assurance that all systematic differences between sites could, in fact, be observed and measured. This latter problem, however, can be partially mitigated by placing the sites themselves into matched pairs and then randomly assigning one member of each pair to treatment status and the other to comparison status. This approach was used for the evaluation of West Virginia's AFDC-U CWEP demonstration and Washington state's Intensive Applicant Employment Services demonstration. Its limitation is that relatively few sites may be available for random assignment. In West Virginia, for example, there were only four pairs of matched sites. Thus, complete randomization may not occur.

Although the advantages of the randomized experimental design over the treatment-comparison site design are apparent, it is important to recognize that the latter does offer certain potential advantages of its own. Specifically, a demonstration may have site-wide effects that will be missed by an impact analysis that relies on an experimental design, but may be captured by an analysis based on a treatment-comparison site design. For example, the very existence of a demonstration at a site may cause workers administering the AFDC program to change their attitudes toward all applicants and recipients, both those in and those outside the target population. Given the presence of a large-scale WEP, for instance, they might begin to administer AFDC regulations more stringently. Moreover, it is also possible that the very existence of a demonstration program at a site, especially one with a CWEP component, may deter some persons who would otherwise be eligible to receive AFDC benefits from applying. Alternatively, a program that stresses training and education should attract persons into AFDC. Since these effects are site-wide, they can only be captured by an evaluation design that compares demonstration sites with sites where the demonstration program has not been implemented.

Another alternative to a randomized experimental design is to compare persons who have elected to participate in a treatment program with persons at the same sites who have elected not to participate. Although this design has been widely used in the past to evaluate human resource programs, it has been used in only one of the OBRA evaluations--an evaluation of the Massachusetts ET-Choices program conducted by the Urban Institute. The ET program is statewide and permits AFDC recipients to voluntarily choose which services, if any, they wish to receive. The problem with this approach is that "observations" in such a demonstration are for persons who selected their own treatment. Hence, members of the treatment group may differ from those in the control in systematic ways that cannot be directly measured--for example in terms of motivation. Recently, considerable professional effort has been devoted to development of methods of unbiased statistical inference in the presence of selection problems (Heckman and Hotz, 1989). The Urban Institute attempted to use such methods in its studies of the Massachusetts ET-Choices program.

Yet another evaluation design that has been used in the OBRA demonstrations is to compare actual outcomes to what would have been expected on the basis of past experience. In its simplest and probably least reliable form, this amounts to attributing all of the change in, say, earnings or welfare payments between the date of the experimental observation and the date used for reference to the demonstration. This is the approach used in evaluations of Oregon's WIN Waiver Jobs Program and Utah's Emergency Work and Employment Program. The problem is, of course, that in the world of public assistance things change even in the absence of a demonstration. A more sophisticated approach is to develop statistical models capable of controlling for the influence of such developments and then evaluating the consequences of experimental developments on the basis of the difference between actual outcomes of interest (dependency, for example) and those which the models would have predicted given all other developments <u>except</u> for the intervention. This is the approach used in a second evaluation of the Massachusett's ET program, one recently completed by June O'Neill (1990) that extends earlier work by Garasky (1990). Forecast-based approaches to program evaluation are viewed skeptically by many, however, in part because of questions about whether or not past experience provides adequate control in the context of rapidly changing economic and social circumstances and because of less-than-successful attempts to use such models to forecast one key outcome measure, welfare caseloads, in the past.[6]

Self-Selection of Sites

Two types of site self-selection occurred in the OBRA demonstrations: interstate and intrastate. Interstate selection occurs because, as described in Section I, under the OBRA regulations, individual states could decide whether or not to conduct a demonstration and, if they did so decide, they had considerable leeway in determining what sort of treatment to offer and what kind of evaluation to conduct. Presumably, each of the demonstration states tended to select treatment combinations for which sufficient political and administrative enthusiasm could be mustered and which appeared administratively feasible.[7] Thus, in interpreting findings from any particular demonstration, it is important to keep in mind that if the same treatment combination was transferred to another state, the results could be very different, particularly if the treatment was imposed involuntarily.

Intrastate site self-selection results because, in most of the demonstration states, the treatment was not implemented on a statewide basis. Instead, only a subset of local offices became demonstration sites. While site selection is generally poorly documented, it appears to have been common to pick locations thought to have compatible and motivated administration. Consequently, findings from demonstrations in these sites may not be generalizable to the rest of the state. There are two reasons for this. First, the demonstration sites may not be very representative of the state as a whole. In the case of North Carolina's CWEP demonstration program, for example, the demonstration sites were six small, rural counties that together accounted for only 7 percent of the state's total caseload. Similarly, the evaluation of Arkansas' WORK program was limited to only two local offices, although one of these covered part of Little Rock. A second reason why it may be difficult to generalize findings from the demonstration sites to the rest of the state is that sites that are self-selected may be better equipped for or more enthusiastic about administering the demonstration treatment than other sites. For example, even in the case of Virginia's Employment Services program, where there is good information on how the sites were selected and considerable effort was made to select sites that as a group were representative of the entire state, the demonstration sites had "all expressed a strong interest in taking part in the study" (MDRC: Riccio et al., 1986, p. 9).

Length of the Study Period

Typically, the period of time covered by the OBRA evaluations was relatively short, and for good reason. Agencies running demonstration programs are understandably anxious to learn of the impacts of these programs. Moreover, the longer an evaluation is continued, the more expensive it becomes. Thus, to cite just one fairly typical example, MDRC began to draw its research sample for its evaluation of the San Diego Job Search and Work Experience demonstration two months after the demonstration began and completed this process eleven months later (MDRC: Goldman et al., 1986, p. 249). Follow-up information was collected for 2 years on the members of the sample selected earliest and for 1 1/2 years for members of the sample selected last.

The fact that the evaluation of the OBRA demonstrations usually covered a period of time close to when the treatment was first implemented may cause the impact measured by the evaluation to understate the ultimate impact of the treatment. It takes time to implement a new program fully. Institutional adjustments may have to be made at administrative agencies, and a learning curve for efficiently running the program may exist. For example, the proportion of later enrollees in Arkansas' WORK program that actually participated in program activities was about 50 percent higher than the proportion of earlier enrollees. One explanation for this appears to be that after the demonstration had been in operation for over a year, the Arkansas Department of Human Services issued new program guidelines and procedures intended to increase participation rates (MDRC: Friedlander et al., 1985, pp. 69-71). As will be discussed later, participation rates may, in turn, affect program impacts.

The relative shortness of the follow-up period is also important because it means that it may be difficult to determine how the demonstration treatment affects program participants over the longer run. For example, if six months into the follow-up period members of a treatment group are found to have higher earnings than members of a comparison group, it would be helpful to know whether this gap will tend to grow or diminish over time. If only a year or two remains of the follow-up period, it may not be possible to make a very clear determination. The length of the follow-up period may be even more important for determining program impacts on welfare status, since such impacts often considerably lag increases in earnings.

Drawing a Representative Research Sample

Most of the OBRA demonstrations were pilot studies that limited treatment to only a subset of a state's total AFDC caseload. It is obviously important that this subset of cases be reasonably representative of the demonstration program's ultimate target population. For example, as suggested earlier, when a demonstration is limited to only a few of a state's welfare offices, the demonstration sites should be as similar as possible to sites in the rest of the state.

A somewhat similar issue is suggested by the fact that evaluations of several of the demonstrations programs--for example, the San Diego Job Search and Work Experience demonstration and the Washington state Intensive Applicant Employment Services demonstration--were limited to only new AFDC applicants. There is nothing wrong with doing this, but if policymakers later become interested in extending the demonstration treatment to the ongoing caseload--a group that differs in many important respects from new applicants--it is important to recognize that any attempt to use evaluation results based on applicants to predict impacts on current AFDC recipients is highly suspect.

Even when new applicants and prior recipients are both included in the research sample, which was the case for most of the OBRA evaluations, certain issues arise. These issues, which are pertinent when applicants and recipients are analyzed together, but not when a separate analysis is conducted for each group, are most easily seen in the case of an evaluation that is based on a randomly drawn sample of new applicants and a randomly drawn sample of prior recipients. To measure demonstration impacts on the program's target population, it is important that the sampled applicants and recipients together constitute a representative cross-section of that population. However, the sampling procedure that best accomplishes this is not entirely obvious. One possibility, for example, is to sample equal proportions of all new applicants during a particular <u>month</u> and all cases already on the rolls at the beginning of that month. An alternative possibility is to sample equal proportions of all new applicants during a particular <u>year</u> and all cases that were already on the rolls at the beginning of the year. The sample resulting from the second procedure would clearly include a much larger proportion of new applicants than the sample resulting from the first procedure.

The approach actually followed by MDRC in drawing research samples for the evaluations they conducted much more closely approximated the second of the procedures described above than the first. But it is not evident that a carefully considered choice was made between the two alternatives. In the case of West Virginia's AFDC Community Work Experience demonstration, for example, assignment to the research sample began in July 1983 and ended in April 1984, a period of ten-months. The study design called for assigning to the research sample all new applicants over this ten-month period who were eligible for treatment and all persons receiving AFDC prior to July 1983 who were eligible for treatment. A roughly equal number of new applicants were assigned each month, generally soon after they applied for AFDC benefits. However, assignment of the prior registrants took some time. In fact, almost half of the eventual sample of prior recipients had not yet been assigned by the end of September, three months after the assignment process began (MDRC: Friedlander et al., 1986, pp. 56-58). Since some prior registrants dropped off the AFDC rolls before being assigned, they were excluded from the research sample. Consequently, the research sample became more heavily weighted towards new applicants than perhaps intended by the research design.

Program Participation

Earlier in this section, we indicated that in most of the OBRA demonstrations, many AFDC recipients were excluded from the WEP target population. Although the excluded groups varied considerably among the demonstrations, they typically included single parents with young children, persons with language barriers or health problems, and persons who lived in inaccessible rural areas.

As described earlier, for purposes of most of the impact analyses that we reviewed, a research sample was drawn that was intended to be approximately representative of the demonstration target population. Inevitably, researchers had room to exercise a considerable amount of judgment in drawing these samples; different choices here and there would have meant different conclusions. For example, the research sample used in the Arkansas WORK program evaluation consisted of 1,153 individuals. However, an additional 940 individuals, who otherwise qualified for treatment under the program, were placed in a "supplementary sample," which was analyzed separately on the grounds that persons in this sample had serious employment barriers and, consequently, had a lower priority for services than members of the main research sample (MDRC: Friedlander et al., 1985, p. 21). Although members of the supplementary sample were less than half as likely as members of the main sample to receive services, they did, in fact, receive some services (MDRC: Friedlander et al., 1985, p. 155). Treatment impacts on the supplementary sample were only reported in an appendix of the evaluation report. For reasons to be discussed shortly, had these individuals been included in the main research sample, the impact estimates for the Arkansas demonstration would have been substantially lower than those reported in the executive summary and main text of the evaluation report.

Once the research samples were selected, members of these samples were allocated to comparison and treatment groups, sometimes randomly and sometimes according to the discretion of program administrators. However, as our process analysis indicates, substantial proportions of persons in the treatment groups typically were not actually treated; that is, they did not actively participate in the demonstration program for which they were eligible. For example, although participation in the OBRA demonstrations conducted in San Diego, Baltimore, Virginia, Arkansas, and Cook County was supposedly mandatory, the percentage of persons in the five treatment samples who actually ever participated in any program activity ranged from a high of 58 percent in Virginia to a low of 38 percent in Arkansas (Friedlander, 1988, p. 19).

There are many reasons why participation rates fall well below 100 percent. Probably the most important is that members of the treatment group drop out of AFDC prior to being required to participate, undoubtedly, in some instances, to avoid participating. In addition, sanctioning may be ineffective or weakly enforced, budgets for WEP activities may be constrained, there may be administrative impediments to more full participation, and some persons may be excused from participating because of ill health or family responsibilities, for example.

Given treatment and comparison groups, estimates of impacts are made by comparing these two groups. Because the treatment group contains both program participants and nonparticipants, however, the impact estimates obtained from this procedure are a weighted average of program effects on persons who received the services incorporated into the treatment and persons who did not. Given this, one would anticipate that the magnitude of impact estimates from the OBRA demonstrations are partially driven by the magnitude of demonstration participation rates; in general, the larger the latter, the larger the former. Consequently, it should be of interest to measure demonstration program impacts in terms of their effects on only those who were actually treated under the program, especially if participation rates can be manipulated by changes in policy. There has been little attempt in the OBRA evaluations to do this, however, because of the substantial statistical problems involved in measuring treatment effects for participants when participation is itself an experimental outcome.

To see this, let us examine two alternative approaches that might be taken. In the simplest of the two, program impacts are estimated in the standard way, by comparing the entire treatment group--those who did not receive the treatment, as well as those who did--with the comparison group, and then divided by the proportion of the treatment group that actually received the program treatment. For example, MDRC estimated that impacts of the Virginia and Arkansas demonstrations on the average earnings of treatment group members were $72 and $70, respectively (Friedlander, 1988, p. 58). Using the procedure just outlined, these measured impacts increase to $124 ($72/.58) and $184 ($70/.38), respectively. Although the latter pair of figures is certainly more impressive than the former, the procedure used to compute the larger estimates has a potentially serious flaw; it is based on the rather tenuous assumption that the demonstration programs affected only those who were active in them. This is unlikely to be true, especially in the case of demonstrations that included a CWEP component. For example, it seems likely, as already suggested, that at least some nonparticipants dropped off the welfare rolls in order to avoid participating. Thus, the procedure just described is likely to produce upwardly biased estimates of the magnitude of demonstration impacts on program participants. The size of this bias obviously depends on the extent to which the demonstrations influenced the behavior of nonparticipants.

An alternative approach towards isolating the effects of the demonstrations on those who participated is to compare comparison group members with a subset of treatment group members made up of only those persons who actually received the treatment. Unfortunately, without invoking special statistical procedures, such a comparison is inappropriate. This is most easily seen for demonstrations in which there was random assignment. In these demonstrations, the characteristics of members of the treatment group should, on average, be very similar to those of members of the comparison group. However, the characteristics of that subset of the treatment group that was active in the program are likely to differ systematically from those of the comparison group. The subset, for example, may have been drawn from among those in the treatment group whom program administrators thought were most likely to profit from receiving the treatment or from among those who were least resistant to participating. Thus, on average, these persons may have had more labor market potential or have been more motivated than members of the comparison group. Although statistical procedures exist to treat such selectivity biases--for example, those mentioned earlier that were developed by Heckman and Hotz--the success of these techniques in any specific instance is somewhat problematic.

The fact that the impact measures produced by most of the OBRA evaluations are weighted averages of program effects on participants and nonparticipants raises several important issues in interpreting the evaluation findings. One approach towards examining these issues is to define a given amount of a specific type of treatment under a particular demonstration program--say, for example, one week in a CWEP program--as a "treatment unit." In the OBRA demonstrations, the number of treatment units varied among members of the treatment group. Nonparticipants, of course, received zero units; but participants themselves received different service quantities. For example, some remained in CWEP longer or participated more fully in job search than others. Moreover, as indicated by Table 1, in many of the OBRA demonstrations, the types of services received varied among participants. For example, some participated in individual job search, some in group job search, some in CWEP, and some in vocational training, and still others participated in several of these components.

In most of the evaluation reports we reviewed, information is provided on the extent of program participation. As discussed in Section III, MDRC, in particular, provides a considerable volume of data concerning participation patterns. It is important that impact estimates be interpreted in light of this information. For had the quantity and types of services received by the treatment population been different, the impact estimates undoubtedly would have also differed.

This issue becomes especially important if an impact analysis is used to predict what would happen if a demonstration program were to become permanent. Such a prediction will become increasingly less accurate, the greater the extent to which the demonstration and the permanent programs differ from one another in terms of any of the three following variables:

- Program participation rates.
- Quantity of program services received per individual, given participation.
- The mix of service types received by each participant.

It may be instructive to discuss a simple example briefly Assume that, as compared to a demonstration, participation rates in a permanent program would be substantially higher; but the other two variables listed above would remain constant. Under such circumstances, it seems likely that any positive program impacts that were found in evaluating the demonstration would be larger in the permanent program. However, exactly how much larger is difficult to predict with precision since it would depend upon such factors as whether scale effects exist, whether program resources must be stretched thinner as the participation rate increases, and whether the new participants would have characteristics that differ from those of the old participants. Predictions obviously become more difficult if two of the variables, rather than just one, would differ, especially if they would differ in opposite directions--for example, if participation rates would increase under the permanent program, but the quantity of services received per participant would decrease. Since actual permanent programs are likely to differ from demonstration programs in terms of all three of the participation variables listed above, not just one or two, generalizing from a demonstration program to a permanent program should be done only with great caution.

We now turn to a somewhat different implication of the fact that in the OBRA demonstrations the quantity and type of services received varied across members of the treatment groups. As mentioned earlier, considerable effort has been made in some of the OBRA evaluations to determine whether treatment impacts vary across subgroups--for example, applicants and recipients, persons with and without lengthy welfare histories, persons with and without previous work experience, and cases that are located in different types of geographic areas. When impact differences are found among subgroups, however, it is difficult to know exactly what to make of them. The subgroups obviously differed in terms of personal characteristics; but, even within the same demonstration, they typically also differed in terms of the three participation variables listed above. These latter differences may, in turn, have been necessitated by the differences in personal characteristics or, alternatively, they may have reflected policy and administrative choices. Hence, one cannot be certain whether an estimated difference in demonstration impacts between two subgroups is attributable to differences in personal capability or receptivity between the two or to administrative factors that resulted in differences in the way they were treated under the program.

Comparisons of Alternative Treatments

As Table 1 suggests, the states that conducted the OBRA demonstrations experimented with a wide variety of treatments and treatment combinations. Determining which of these treatments or combinations of treatments tended to produce the greatest impacts is obviously of great policy importance. There are two potential means of doing this: comparisons across demonstrations and comparisons within demonstrations.

The first of these approaches exploits the fact that the OBRA demonstrations did vary considerably in the treatments they offered. However, although comparisons of impact estimates for different demonstrations can be instructive, they should be interpreted with considerable caution, for in addition to treatment variation, there were other important differences among the demonstrations.

For example, as suggested in the previous subsection, program participation rates varied considerably. Moreover, the states conducting the demonstrations differed in terms of labor market conditions, population density, AFDC payment levels, characteristics of the welfare population, and approaches toward administering welfare. It is clearly difficult to untangle all these factors in order to isolate the influence of treatment variation on observed differences in program impact. Indeed, sometimes the environment in which a demonstration is conducted may tend to dominate program outcomes. For example, different variants of CWEP were tested in San Diego and West Virginia. Since San Diego's unemployment rate first peaked at about 10 percent and then fell to 6 percent during the demonstration study period, while West Virginia's unemployment rate peaked at 18 percent and then declined to only 13 percent, it is not too surprising that earnings impacts were found to be substantial in San Diego, but nonexistent in West Virginia.

The second means of comparing the impacts of alternative treatments, comparisons within demonstrations, is potentially superior to the first. This approach requires that there be two or more treatment groups, rather than just one, with each group receiving a different treatment. The treatment groups can be compared to one another, as well as to a comparison group. Thus, it is possible to determine which treatment had the larger impact. Unfortunately, the research design just described was utilized relatively rarely in the OBRA demonstrations, and in only one or two instances were impact estimates produced that are useful in comparing alternative treatments.

It may be instructive to examine briefly one demonstration in which the multiple treatment group design was used successfully and one in which it was not. The successful example is the San Diego Job Search and Work Experience demonstration, which utilized two treatment groups. The first group was required to participate in three weeks of supervised job search activities. The second group was also required to participate in three weeks of job search activities but, in addition, those individuals who had not found jobs by the end of the three weeks were required to hold work experience jobs for up to thirteen weeks. Thus, this design permitted measurement of the incremental impact of CWEP.

A somewhat similar, although more complex, research design was used in the Virginia Employment Services demonstration, but less successfully. In this case, members of both treatment groups were required to participate in job search activities. In addition, after completing job search, members of one of the treatment groups could be required to participate in a work experience program if Department of Social Services staff thought it appropriate, while members of the second treatment group could be assigned to a work experience program, an educational or training program, or both, if staff thought it warranted. Thus, the intention was to measure the incremental impacts of education and training. Unfortunately, since AFDC recipients in Virginia can enter educational and training programs voluntarily on their own initiative, it turned out that members of the first treatment group were almost as likely to participate in such programs as members of the second group. Indeed, substantial numbers of persons in the control group also voluntarily enrolled in educational and training programs. Because of implementation problems, therefore, little could be learned from the Virginia demonstration concerning the incremental effects of employment and training.

Before leaving our discussion on comparing the impacts of alternative treatments, it may be useful to comment on one potential treatment variation that was essentially ignored in developing the research designs of the OBRA demonstrations: the use of monetary sanctions. Almost all the programs tested in the demonstrations were mandatory, and the use of sanctions to enforce participation in such programs has been widely discussed. The impact of sanctions might have been tested in one of the demonstrations by using a two-treatment-group research design, similar to that described above, and vigorously enforcing strong sanctions on one of the treatment groups, but not the other. This was not done, however. Although there was some variation among demonstrations in the use of sanctions, the demonstrations also varied in so many other respects that it is not possible to draw any conclusions concerning their impact.

Estimation Issues

Level of Aggregation

Most of the OBRA impact analyses used individual AFDC cases as their unit of analysis. A few, however, were based on a more aggregate unit of analysis: individual welfare offices or individual states. Estimating procedures in these two types of studies were quite distinct from one another, and therefore will be discussed separately.

The Aggregate Approach. Impact analyses that use individual welfare offices or states as the unit of analysis are mainly concerned with whether the treatment tested in a demonstration reduces the size of the AFDC caseload. One method for determining this is to utilize a caseload forecasting model to predict what caseload levels would have been in the demonstration sites in the absence of the program and then compare this prediction with actual caseload levels in the presence of the demonstration. If the predicted caseload exceeds the actual caseload, this would imply a caseload reduction attributable to the tested program.

This approach has been used in evaluating both the North Carolina Community Work Experience Program demonstration and the Ohio Work programs. In both of these states, the demonstration program was tested in a small, nonrandomly selected subset of counties. The forecasting models used in the two evaluations were based on data for both demonstration and nondemonstration counties. The forecasting model used in the North Carolina evaluation was an autoregressive time-series model routinely used by the state to predict caseloads in individual counties. The model used in the Ohio evaluation was specifically constructed for purposes of the evaluation and was based on cross-sectional time-series regressions. Both models predicted caseload levels in the demonstration counties that exceeded observed levels, implying that the demonstration treatment reduced caseloads.

Although the evaluation technique just described is a reasonable one, the danger always exists that differences between predicted and observed caseload levels are due to some factor other than the demonstration treatment, a factor that is not adequately captured by the forecasting model. This is of particular concern because, as mentioned, the demonstration sites in North Carolina and Ohio were not randomly selected.

For two different reasons, the estimated caseload impacts for the North Carolina demonstration appear superior to those for the Ohio demonstration. First, the North Carolina forecasting model was based entirely on data for a time period that preceded the demonstration. The Ohio model, in contrast, relied on data that covered the period while the demonstration was in progress, as well as data from the period prior to initiation of the demonstration. Thus, to an unknown extent, the caseload forecasts for the demonstration counties in Ohio were influenced by the impact of the treatment being tested. Consequently, these forecasts are an imperfect predictor of what caseload levels in the demonstration counties would have been in the absence of the treatment. Second, the North Carolina evaluation report presented caseload forecasts for the nondemonstration sites, as well as for the demonstration sites. Consequently, it is possible to adjust the North Carolina impact estimates for forecasting errors that affected the state as a whole. The information needed to make such adjustments is not provided for the Ohio evaluation.

An alternative method to the one just discussed for determining demonstration impacts on aggregate caseloads is to compare caseload levels at demonstration sites with those at nondemonstration sites, using regression analysis to control for other differences between the two types of sites. In such an analysis, estimated differences in caseload size are attributed to the impact of the demonstration. This approach was rather convincingly utilized by O'Neill (1990) in a study that used national data to attempt to determine whether the Massachusetts ET program reduced that state's AFDC caseload level. It has also been utilized, but less convincingly, by Bradley R. Schiller and C. Nielson Brasher (n.d.) in an attempt to use national data to determine whether states with CWEP programs have smaller caseloads, everything else being equal, than states without such programs.

The basic regression model used by Schiller and Brasher, which is based on quarterly data for each state over the period from 1981 to 1986, can be represented as follows:

$$CLR(i,t) = a + b \cdot CWEP(i,t) + c \cdot S(i) + d \cdot X(i,t) + f \cdot Q(t) + e,$$

where CLR(i,t) is AFDC caseload divided by population in the ith state during the tth calendar quarter; CWEP(i,t) is a dummy variable that equals one if the ith state was operating a CWEP program during the tth quarter and zero otherwise; S(i) is a vector of dummy variables representing all but one state; X is a vector of control variables intended to capture conditions unique to each state (for example, each state's labor force participation rate, poverty rate, population density, maximum AFDC benefit, and AFDC standard of need); Q(t) is a vector of dummy variables representing all but one of the calendar quarters covered by the data; e is an error term; and a, b, c, d, and f are parameters to be estimated.

Given this model, the estimate of b is used by Schiller and Brasher as a measure of the impact of CWEP on state caseload levels. In assessing this model, at least three points can be made. The first is a standard issue: in the absence of random assignment, one cannot be certain that the regression sufficiently controlled for factors other than CWEP that may have influenced differences in caseload levels between states with and without CWEP. For example, the value of CLR in 1980 might have been included as an additional control variable in the regression. Or, alternatively, a longer time series, one that included the period prior to the passage of OBRA and the establishment of CWEP programs, might have been used. Second, given the regression specification, the vector of dummy variables represented by S(i) is probably highly collinear with the CWEP dummy. This follows from the fact that some states had CWEP programs in place throughout most of the 1980s while many had none at all over this period. Consequently, the estimate of b probably does not measure the impact of CWEP very precisely. Third, "having CWEP" can mean many different things, ranging from a one- or two-county demonstration to a statewide implementation. As we discussed in connection with workfare process, at minimum it would seem important to assess the incidence of CWEP among potential participants to see just how much work experience the states in fact offered.

The Disaggregate Approach. As already noted, most of the OBRA impact analyses utilized information on individual cases within experimental and comparison groups. In some of these analyses, impacts were measured by simply computing the mean values of the outcome measures of interest for each research group and then comparing these means without any test of statistical significance. More typically, however, ordinary least squares regression equations were estimated, and tests of statistical significance were made. The basic regression model used for this purpose can be represented as follows:

$$Y(i) = a + b \cdot T(i) + c \cdot Z(i) + u,$$

where Y(i) is an outcome measure for the ith case (for example, welfare status, AFDC receipts, employment status, or earnings); T(i) is a dummy variable that equals one if the case is in the treatment group and zero if it is in the comparison group; Z(i) is a vector of personal characteristic variables used for control purposes (for example, measures of age, education, work experience, time on AFDC, marital status, number of children, and race); u is an error term; and a, b, and c are parameters to be estimated. The estimate of the coefficient b, which may be viewed as the adjusted mean difference in outcome between the treatment and comparison groups, provides a measure of demonstration impact.

The use of a regression model to estimate demonstration impacts, rather than a simple comparison of means, has two advantages. First, it increases statistical efficiency. Second, and more importantly, it helps correct for predemonstration differences between members of the treatment and comparison groups that may affect the outcomes of interest. Such corrections are likely to be minor if random assignment was used in allocating cases to treatment and comparison groups, but may be crucial in the absence of random assignment. Somewhat ironically, in the OBRA evaluations, regression analysis was more likely to be used when random assignment was present than when it was not.

When regression analysis has been used in the OBRA evaluations, there has been almost exclusive reliance on ordinary least squares (OLS) techniques. Although OLS regressions are inexpensive, convenient to use, and easy to interpret, they are not necessarily appropriate for the OBRA impact analyses. The reason for this is that all the dependent variables in these analyses--for example, dummies indicating whether or not welfare or earnings were received over a given time interval and continuous variables indicating amounts of welfare or earnings received (including zero values)--were bounded, either at both ends or at the lower end. As a consequence, the error terms in these models will be neither normally distributed nor homoscedastic. The lack of homoscedasticity may be particularly problematic, since the standard errors reported by conventional statistical packages are based on the assumption of homoscedastic errors. The true standard errors would be larger, perhaps considerably larger. The possibility that these issues may be important is suggested by the fact that in the MDRC evaluations the dependent variables were usually measured over individual calendar quarters and, during any given quarter, from 60 to 80 percent of the research samples used in the various studies were typically not employed and from 30 to 60 percent typically did not receive any AFDC benefits.

It would appear that sensitivity tests in this area are warranted to determine how important the issues just discussed actually are. For example, in the case of dependent variables bounded at one end, such as earnings and AFDC receipts, procedures are available that compute correct or robust standard errors that (even in the presence of heteroscedasticity) could be used, and the resulting findings could then be compared to those from uncorrected OLS estimates. In addition, hypotheses tests that are based on more appropriate t-value tables (see White, 1980) could be compared to those based on standard tables. A similar approach could be followed in the case of zero-one dependent variables, for example employment and welfare status, where t-values from probit or logit regression estimates could be compared to those from OLS estimates.

As the very simple reduced-form regression equation that appears above suggests, even the more sophisticated OBRA impact analyses were conducted with virtually no reference to economic theory. This was true in spite of the extensive development of human capital theory and job search theory during the past two decades and despite the fact that many of the programs tested were mainly concerned with increasing the employability of AFDC recipients and placing them into unsubsidized jobs. The fact that the OBRA impact analyses have been little influenced by economic theory has several implications, which we now briefly discuss.

First, although it would not appear difficult to make theoretical arguments that Job Clubs would be expected to increase the probability of holding a job, that vocational training should increase earnings, and that CWEP should encourage AFDC recipients to leave the rolls, tests of statistical significance in the OBRA demonstrations were almost inevitably two-tailed. This seems overly conservative.

Second, because structural models were not estimated in the OBRA evaluations, impact estimates from any given demonstration should not, in a strict sense, be generalized to population groups that differ from the research sample used in the demonstration--for example, groups with greater or less prior work experience or education. Of course, it is theoretically possible to break the research sample into subgroups and obtain separate impact estimates for each; but, as discussed earlier, interpreting findings from such an exercise is difficult, even if the data sets are sufficiently large to permit grouping of the data.

Third, findings from the evaluations that may appear surprising to many readers typically received ad hoc explanations or no explanation at all; there were simply no theoretical predictions to which to compare them. For example, although the treatment under the Arkansas WORK program consisted almost entirely of individual and group job search, estimated program impacts were not only found to persist over twelve calendar quarters, but those on earnings seemed to grow (Friedlander and Goldman, 1988, Table 1). It seems likely that, given a theoretical model, such a finding would have been unanticipated. A simple conceptual model, for example, might have predicted that supervised job search would give those who received it a head start in obtaining employment over those who did not, but that the latter would soon catch up with the former. Nevertheless, no attempt was made to explain the finding.

Data Sources

In general, the data required to conduct the OBRA impact evaluations were readily obtained. The data necessary to measure the independent variables needed for the regression model described in the preceding subsection all pertain to the characteristics of individuals at the time of their assignment to the research sample. Since these persons were either receiving or applying for AFDC benefits at this time, the required data could be easily obtained directly from them. In fact, much of the necessary information was routinely collected in administering the welfare system.

Measuring the dependent variables required information on individuals after their initial assignment to the research sample. The data needed to measure two of these variables--welfare status and amount of welfare received--could of course be readily obtained from the administrative records of the agencies conducting the demonstrations. Two alternative sources of data were available to measure the remaining dependent variables--employment status and earnings.

The first of these data sources is follow-up interviews with members of the treatment and comparison groups. This approach is expensive and time consuming and it leads, if those who cannot be found or who will not participate in interviews differ in earnings or employment from those who do participate, to "attrition bias." This problem would occur, for example, if the treatment provided by a demonstration program helps members of the treatment group secure employment and, thereby, leave the welfare rolls, and it is easier to obtain interviews with those still on the rolls than with those who are off. Under such circumstances, the impact of the demonstration treatment on employment would be understated.

The second source for employment and earnings data is state Unemployment Insurance (UI) earnings records. These records are constructed from quarterly earnings reports that are required from employers for purposes of administering state UI programs. In general, the UI earnings data are preferable to interview-based earnings data, since UI data are less expensive to obtain and are less subject to attrition bias. They may also be more accurate, since they are not subject to recall errors. However, although all states are now required to maintain UI earnings files, some states did not have these data bases during the time period covered by the OBRA evaluations or the managing agencies were unwilling to cooperate in the data collection effort. Consequently, interview-based earnings data were used instead of the preferable UI earnings data in conducting several of the OBRA impact analyses.

Although UI earnings records were the best source of data on earnings and employment status for use in the OBRA demonstrations, they did suffer from one important limitation: they were not available for all employed members of the research samples. There are several reasons for this. First, coverage is incomplete. Excluded groups include federal employees; railroad workers; the self-employed; some domestics; some farm workers; and, most important, individuals employed in the underground economy. Second, data were not available for research sample members who found employment in other states, either because of a geographic move or because they lived near a state border. Third, obtaining UI earnings records on research sample members requires a computer match between social security numbers maintained in the administrative records of welfare offices and social security numbers reported by employers. Any errors in either social security number for a given individual would cause a mismatch. This mismatch can go in either direction: earnings may be assigned to an individual who is not working, or the earnings of someone who is working may be missed.

The first two of the factors listed above will cause the earnings and employment status of members of research samples to be underreported; the third factor will cause them to be misreported, with the direction of error uncertain. However, because this underreporting and misreporting should not, in general, differ between treatment and comparison group members, any biases to impact estimates that rely on UI earnings data will probably be small. Nevertheless, two possible sources of bias, both of which would tend to understate earnings impacts, do exist. The first of these, which appears likely to be minor, would occur if some members of a treatment group were to move to another state in order to avoid mandatory participation in a program such as CWEP and later obtain employment in their new state. The second bias is most easily described through use of a numeric illustration. Assume that, according to information provided by UI earnings data, one year after being assigned to a research sample, 40 percent of the members of the treatment group are employed, but only 30 percent of the comparison group work, thus, implying that the demonstration treatment has increased employment by 10 percentage points. Assume further that because of underreporting, actual employment levels are understated by 20 percent. In other words, 48 percent of the treatment group and 36 percent of the comparison group actually work. As is apparent, this implies that because of underreporting, the estimate of the demonstration impact, which was based on UI earnings data, understated the true impact of the demonstration by 2 percentage points.

Use of earnings data from the UI system creates one additional anomaly in evaluation of OBRA impacts worth noting. UI reporting is done on the basis of actual calendar quarters. Most other information utilized in OBRA evaluations is developed on a monthly basis. Most evaluations translate data derived from monthly reporting back to calendar quarters to conform to the earnings information. The result often appears to be information defined relative to, say, the point of application or welfare acceptance. But in this environment "quarter 1" does not mean the first three months after, say, beginning a WEP. Instead "quarter 1" means the calendar quarter that includes the point at which WEP participation was initiated. Thus, an applicant for public assistance in December may be assigned to a treatment group and be approved for welfare. But if her first check is paid in January, she will show no welfare receipt until quarter 2, and what is termed quarter 1 may include earnings from employment before welfare application. This timing problem suggests that in evaluating UI system-based data on OBRA effects, one should ignore quarter 1.

Findings from the Impact Analyses

Of the twenty-four OBRA demonstration evaluations listed in Table 1, impact analyses had been completed for nineteen by the end of 1989. We now turn to their findings. In doing this, we limit our discussion to only thirteen of the available nineteen sets of findings--those in which we place some confidence. Ten of the thirteen sets of findings we discuss were obtained from evaluations conducted by MDRC. The remaining three are from evaluations of Massachusetts' Employment and Training (ET) Choices program, North Carolina's CWEP demonstration program, and Washington state's Intensive Applicant Employment Services project. The first of these three evaluations was conducted by June O'Neill, an academic. The other two were conducted in-house by state agency personnel.

Before proceeding, a few comments are in order about the six excluded evaluations: Ohio's Work Experience Program, Minnesota's Work Experience Program, Oregon's WIN Waiver Jobs Program, Utah's Emergency Work and Employment Program, Florida's TRADE Welfare for Work program, and Washington state's Community Work Experience Program.

Our lack of confidence in impact findings from these evaluations stems from a variety of causes. As discussed earlier in this section, findings from the Ohio Work Experience impact analysis were subject to important econometric problems. Impact findings from the Minnesota, Oregon, and Utah demonstrations are based on relatively crude comparisons between two sets of sites, those with and without the treatment, or between two points in time, prior to and after introduction of the treatment. These comparisons are subject to compounding differences between the two sets of sites or the two points in time that are inadequately controlled for by the analyses. Indeed, in two of the three evaluations, no attempt at all was made to control for these factors. Consequently, reliable inferences as to impacts cannot be drawn. The Florida and Washington state demonstrations did utilize an experimental design. Unfortunately, after random assignment, numerous cases were dropped from the samples selected in both demonstrations, in large part because of missing data. Since the cases that were dropped differed systematically between the treatment and comparison groups, the integrity of the random assignment was violated and, as a consequence, the impact findings, which are based on comparisons between the treatment and comparison groups, are unreliable.

We now consider findings from the thirteen impact analyses in which we do place some confidence. We have attempted to summarize many of the more salient results from the ten analyses that were conducted by MDRC in Tables 6 through 8. Results from the three non-MDRC evaluations do not lend themselves to the format used in these tables, but will be briefly described later.

The intent in Tables 6 through 8 is to restrict the volume of reported numbers to a manageable level, yet convey important information from the MDRC impact analyses. Thus, while

MDRC uses four measures of impacts, we focus only upon earnings (in Table 6) and welfare status (in Table 7). These two measures would appear to capture adequately how well the OBRA demonstrations met the major objective of increasing the earnings of AFDC recipients, thereby allowing them to leave the welfare rolls. Parentheses are used to indicate estimates of treatment impacts that came out the opposite from this goal--that is when earnings were found to decrease or persons remaining on AFDC were found to increase.

Tables 6 and 7 contain five identical columns. The first column reports the control group mean of the impact measure. The intent here is to provide a benchmark, an estimate of the value of the impact measure in the absence of the treatment. The second column presents MDRC's estimates of the impacts themselves. Based on the fact that in its reports MDRC provides estimates that pertain to individual calendar quarters, the numbers found in the first two columns are averages that are computed over several quarters. In making these computations, we began with the third calendar quarter after an individual was assigned to the evaluation sample and included every additional quarter thereafter for which impact estimates existed. The first two quarters were not used in the computations because of the ambiguity of first-quarter definition and because services and activities for most of the demonstrations extended at least through quarter 2. The third and fourth columns report impact estimates for two individual calendar quarters: the third and the last available. These estimates can be compared to one another or, better, with the average impact estimates in the second column to obtain a rough idea of how the impact of each demonstration treatment varied over time. The final column in each table indicates the total number of calendar quarters over which the averages were computed and the number of these quarters for which the impact estimates were statistically significant, at least at the 10 percent level. Thus, for example, the average impact on earnings estimate that appears in the first row of the second column of Table 6 is based on individual impact estimates for the last four of the six calendar quarters for which MDRC provided estimates. And of

TABLE 6

Summary of Earnings Effects Found in MDRC Evaluations of Various OBRA Demonstrations (Earnings received in a calendar quarter-year, in current dollars)

Experimental Group and Demonstration Location	Control Group: Average Earnings in Absence of Program	Program Impact: (Treatment Group Minus Control Group) Average Impact[a]	Impact in Third Quarter	Impact in Last Quarter	Significant Quarters among Available Quarters[b]
AFDC-R applicants					
San Diego EPP/EWEP:					
Job search only	$ 683	$ 33	$ 118**	$ 26	1/4
Job search/CWEP	683	140	163***	161***	4/4
Baltimore	823	181	94	194	9/10
Arkansas	186	(-6)	(-6)	(-6)	0/1
West Virginia	228	(-40)	(-21)	(-55)	0/4
Virginia	526	78	69	87	0/2
Cook County:					
Job search only	690	(-61)	(-84)*	(-19)	1/4
Job search/CWEP	690	(-48)	(-62)	(-24)	0/4
San Diego SWIM	685	90	76	29	2/7
AFDC-R recipients					
Baltimore	427	40	48	18	0/3
Arkansas	42	42	42**	42**	1/1
West Virginia	124	14	38**	10	1/4
Virginia	273	19	19	20	0/2
Maine	590	207	99	209*	8/9
New Jersey	988	109	40	123	2/5
Cook County:					
Job search only	266	39	18	46*	3/4
Job search/CWEP	266	45	37*	44*	4/4
San Diego SWIM	402	172	87**	221***	7/7
AFDC-U applicants					
San Diego EPP/EWEP:					
Job search only	1,580	59	136	15	0/4
Job search/CWEP	1,580	23	69	7	0/4
San Diego SWIM	1,096	139	183	145	0/7
AFDC-U recipients					
San Diego SWIM	542	91	105	110	0/7
AFDC-U combined					
Baltimore	1,499	(-416)	(-400)	(-551)	2/3
West Virginia	608	(-38)	(-16)	(-102)***	1/4

Source: Various final and supplemental reports on individual state work/welfare demonstrations published by the Manpower Demonstration Research Corporation.

Note: Estimated effects reported in parentheses have unanticipated sign.

[a]Averaged from quarter 3 through last quarter available.

[b]Number of quarters with a statistically significant increase or decrease, from quarter 3 through last quarter available.

* Statistically significant at the 10% level using a 2-tailed t-test.

** Statistically significant at the 5% level using a 2-tailed t-test.

*** Statistically significant at the 1% level using a 2-tailed t-test.

TABLE 7

Summary of Effects on AFDC Receipt Found in MDRC Evaluations of Various OBRA Demonstrations (Percentage receiving AFDC at any time in a calendar quarter-year)

Experimental Group and Demonstration Location	Control Group: Average Receipt in Absence of Program	Program Impact: (Treatment Group Minus Control Group) Average Impact[a]	Impact in Third Quarter	Impact in Last Quarter	Significant Quarters among Available Quarters[b]
AFDC-R applicants					
San Diego EPP/EWEP:					
Job search only	45.9 %	-1.8 %	-6.0 %	(0.0) %	1/4
Job search/CWEP	45.9	-2.8	-6.4**	-1.2	1/4
Baltimore	50.1	-2.4	-1.4	-1.1	0/10
Arkansas	48.5	-3.3	-3.3	-3.3	0/1
West Virginia	51.6	-0.2	(0.8)	-3.5	0/5
Virginia	43.0	-1.7	-1.0	-2.4	0/2
Cook County:					
Job search only	67.9	-4.3	-3.5**	-5.3***	4/4
Job search/CWEP	67.9	(2.4)	(2.6)	-2.1	0/4
San Diego SWIM	52.4	-6.3	-4.5*	-6.9***	8/8
AFDC-R recipients					
Baltimore	74.0	(0.3)	(0.1)	(1.3)	0/10
Arkansas	85.7	-11.4	-11.4***	-11.4	1/1
West Virginia	75.7	-2.5	-1.8	-2.5	2/5
Virginia	77.8	-0.5	-2.3	(1.2)	0/2
Maine	68.6	(1.9)	(0.0)	(6.0)	0/9
New Jersey	60.7	-3.5	-0.9	-1.8	1/6
Cook County:					
Job search only	84.7	-2.3	-0.9	-3.1***	3/4
Job search/CWEP	84.7	-1.3	-1.0	-1.5	0/4
San Diego SWIM	76.8	-6.4	-1.5	-7.0	7/8
AFDC-U applicants					
San Diego EPP/EWEP:					
Job search only	40.4	-4.7	-7.0***	-5.1**	3/4
Job search/CWEP	40.4	-5.0	-7.4***	-2.9	3/4
San Diego SWIM	49.7	-3.2	-1.7	-1.7	1/8
AFDC-U recipients					
San Diego SWIM	73.4	(4.0)	-7.6	-4.6	1/8
AFDC-U combined					
Baltimore	43.0	(4.8)	(4.9)	(2.8)	0/3
West Virginia	59.0	-5.3	-4.8***	-6.9***	5/5

Source: Various final and supplemental reports on individual state work/welfare demonstrations published by the Manpower Demonstration Research Corporation.

Note: Estimated effects reported in parentheses have unanticipated sign.

[a]Averaged from quarter 3 through last quarter available.

[b]Number of quarters with a statistically significant increase or decrease, from quarter 3 through last quarter available.

* Statistically significant at the 10% level using a 2-tailed t-test.

** Statistically significant at the 5% level using a 2-tailed t-test.

*** Statistically significant at the 1% level using a 2-tailed t-test.

these individual four impact estimates, only one was statistically significant at at least the 10 percent level.

The two columns in Table 8 indicate the sample sizes upon which the impact estimates are based and program costs per experimental. The reported sample sizes are based on members of both treatment and control groups. As can be seen, all but a few of the sample sizes are relatively large. However, in analyses of treatment impacts on subgroups, a topic we will discuss shortly, the samples in some of the demonstrations were sometimes stretched rather thin.

The estimates of program costs per experimental were obtained from the cost-benefit sections of various MDRC reports. These figures include estimates of program operating costs, allowances paid to program participants (mainly for day care and transportation expenditures), and, in the case of the New Jersey and Maine demonstrations, wage subsidies paid to private sector employers. The estimates are net of similar expenditures on comparison group members. The purpose of providing these estimates is to try to take account of the fact that participation levels varied widely among various treatment groups, an issue discussed earlier in this section. By measuring the government's average investment in each member of the different treatment groups, the cost estimates provide a rough gauge of the number and type of services received by these persons.

Results in Tables 6 through 8 are reported separately for the AFDC-R and the AFDC-U programs. In addition, whenever possible, impact estimates are reported separately for new program applicants (that is, those designated in Table 1 as APPs, New Recipients, or NDMRs) and prior program recipients (those designated as CMRs). Not only did treatment impacts tend to differ substantially between these two groups, but as the last column of Table 8 suggests, the average cost of treatment did also. For two of the demonstrations--the one in Cook County and the first one in San Diego--there were two separate treatment groups. Impact findings are reported separately for each of these two groups.

TABLE 8

Summary of Program Costs in MDRC Evaluations of Various OBRA Demonstrations

Experimental Group and Demonstration Location	Sample Size	Program Cost per Treatment Group Member[a]
AFDC-R applicants		
San Diego EPP/EWEP:		
Job search only	1,729	$ 510
Job search/CWEP	2,375	578
Baltimore	1,380	815
Arkansas	667	158[b]
West Virginia	1,078	178
Virginia	1,286	253
Cook County:		
Job search only	2,631	127[b]
Job search/CWEP	2,668	158[b]
San Diego SWIM	1,258	700
AFDC-R recipients		
Baltimore	1,377	755
Arkansas	452	158[b]
West Virginia	2,601	301
Virginia	1,897	548
Maine	444	2,286
New Jersey	994	861
Cook County:		
Job search only	5,231	127[b]
Job search/CWEP	5,187	158[b]
San Diego SWIM	1,953	1,068
AFDC-U applicants		
San Diego EPP/EWEP:		
Job search only	1,644	543
Job search/CWEP	2,189	672
San Diego SWIM	798	660
AFDC-U recipients		
San Diego SWIM	543	1,025
AFDC-U combined		
Baltimore	337	537
West Virginia	5,630	136

Source: See Table 1.
[a]Additional government expenditures on treatment group members, relative to expenditures on comparison group members.
[b]Separate program cost estimates for applicants and recipients were not available for Cook County and Arkansas.
* Statistically significant at the 10% level using a 2-tailed t-test.
** Statistically significant at the 5% level using a 2-tailed t-test.
*** Statistically significant at the 1% level using a 2-tailed t-test.

The impact estimates reported for earnings and welfare receipt vary considerably from one demonstration to another. This is not surprising given the variety of treatments offered by the demonstrations, differences in participation rates among the demonstrations, variations in the target groups covered by the demonstrations, and variations in the economic and institutional environments in which the demonstrations were conducted. Moreover, many of the impacts are estimated with little precision and are therefore statistically insignificant. Nevertheless, with only a few exceptions, they are in the hoped-for direction. That is, they imply that the demonstration treatments really did increase earnings and reduce the size of the welfare rolls.

However, most of the impact estimates indicate that the demonstration treatment effects were relatively small. For example, only six of the twenty-four earnings impact estimates appearing in the second column of Table 6 imply that average earnings were raised by over $100 per calendar quarter by the treatment, and only one of these impact estimates exceeded $200 per quarter. Viewed from an annual basis, it would appear that in no case were earnings increased by as much as $1,000 a year. As pointed out earlier, however, these earnings impacts are averaged over those who participated in treatment under the demonstrations and those who did not. They would almost certainly be higher if estimated only for those who actually received the program treatment. Nevertheless, one is left with the impression from Table 6 that, although some of the treatments tested in the OBRA demonstrations did raise the earnings of some program participants, overall, these effects are quite limited. Since most treatment impacts on rates of employment are well under 10 percentage points in any given calendar quarter, it would appear that whatever earnings increments occurred were concentrated among relatively few persons.

Table 7 shows that the demonstration treatments were also successful in reducing the incidence of welfare receipt, but again only to a limited degree. For example, only four of the twenty-four average impact estimates exceeded 5 percentage points and only one, that for AFDC recipients enrolled in Arkansas' WORK program, exceeded 10 percentage points. Moreover, program success in reducing the size of the welfare rolls was substantially smaller than these estimates might at first blush appear to imply. The reason for this is that, as previously pointed out, not all welfare recipients were included in the target populations of any of the demonstration programs. For example, 38 percent of Arkansas' AFDC caseload had children under the age of three and, consequently, was not required to participate in that state's WORK program. Had these persons been included in the evaluation sample, the estimated impact of the treatment on welfare status would have been considerably smaller.

The results for non-MDRC OBRA evaluations are broadly similar to those reported by MDRC. Both Washington state's Intensive Applicant Employment Services project and North Carolina's CWEP demonstration program also apparently reduced the size of the welfare rolls. For example, the Washington state project, which was limited to new AFDC applicants, appeared either to reduce the amount of time new applicants, once accepted, remained on the rolls or reduce the likelihood that these persons ever got on the rolls to begin with. During the ten months after originally applying for AFDC, members of the comparison group received assistance for 3.5 months, on average, while members of the treatment group received assistance for only 3.1 months, on average (Fiedler, 1983, Table B-1). The impact analysis of North Carolina's CWEP demonstration program, which covered both applicants and recipients, implied that the AFDC caseload was around 4 percent lower than it would have been in the absence of the program (North Carolina Department of Human Resources, 1985, Table 13). There is also some evidence that this program had a small positive effect on earnings. On the other hand, from O'Neill's study (1990, chapter 5) it appears that the Massachusetts ET program had little if any effect on the size of that state's AFDC caseload or on the earnings of its AFDC recipients.

The impact data support several additional inferences, which we will now briefly discuss. First, both new AFDC applicants and persons already receiving AFDC were treated under most of the OBRA demonstrations evaluated by MDRC. Tables 6 and 7 suggest that although some of these demonstrations were more successful in treating one group rather than another, a clear pattern does not appear to emerge. As previously mentioned, however, program expenditures per member of the treatment recipient groups considerably exceeded those per member of the applicant groups. Probably the major reason for this is that applicants were more likely than recipients to leave the AFDC rolls early in the treatment process. Indeed, some never became eligible for welfare at all.

Second, several of the OBRA demonstrations evaluated by MDRC covered both the AFDC-R and the AFDC-U populations. With one exception, it does not appear that these demonstrations were more successful in treating one of these population groups than the other. The major exception occurred under the demonstration in Baltimore, where the lack of success in treating AFDC-U participants is striking. The reader should note, however, that results for AFDC-U participants in Baltimore are based on a relatively small sample.

Third, four pairwise comparisons can be made in Tables 6 and 7 that allow one to examine the incremental effect of adding CWEP to job search. Only in the case of AFDC-R applicants in San Diego did this addition appear to have a positive incremental impact. Positive incremental effects are not apparent for either AFDC-U applicants in San Diego or AFDC-R applicants or recipients in Cook County.

Fourth, for reasons discussed in the next section, where we consider cost-benefit analyses, it is of considerable importance to determine whether the impacts of demonstration programs tend to persist or decay over time. Tables 6 and 7 provide mixed evidence concerning this issue. Some of the measured impacts reported in the table do appear to decay, but others seem to persist or even grow over time. It should be emphasized, however, that most of these impacts are measured over relatively few calendar quarters. If program impacts could be observed for a longer period of time, it might be possible to provide more definitive evidence concerning time trends.

Finally, when crudely averaged the results in Tables 6 and 7 reveal an important difference between AFDC-R and AFDC-U impacts. For AFDC-R cases, across all of the experiments examined, the incidence of statistically significant treatment effects on earnings is greater (43/79) than for AFDC receipt (28/91). The opposite is true for AFDC-U, where the OBRA experiments tended to more regularly identify significant consequences for AFDC receipt (13/32) than for earnings (3/29). The reasons for this differential are unclear. One interpretation is that since the duration of AFDC-U cases is generally much shorter than that of AFDC-R cases, the incidence of treatment receipt is lower; as a result, earnings effects for AFDC-U are smaller and more irregular than for AFDC-R. At the same time, the obligation imposed by the treatment activities accelerated closures. It may be significant that eligibility rules for AFDC-U lead to case closure regardless of income if the family's "principal earner" is working more than 100 hours per month. As a result, some types of employment that would not have closed AFDC-R cases require closure, because of the 100 hours rule, for AFDC-U families. Furthermore, sanction procedures differ between the two types of cases. For AFDC-R, the usual WIN sanction leads only to elimination of the noncompliant adult from inclusion in calculation of benefits for the family. For AFDC-U, a noncompliant "principal earner" means that the family as a whole is no longer eligible, since it is, in principle, the involuntary unemployment of that person that qualified the family in the first place.

This minor mystery points to a significant shortcoming in the analysis of OBRA results. Despite the enormous attention given family stability effects of the negative income tax experiments, none of the OBRA analyses looks at the consequences for the stability of dependent two-parent families of the OBRA treatments. Family separation in AFDC-U cases is quite common (see Schram and Wiseman, 1988), and it is readily identifiable by the transition of the AFDC-U case to the

AFDC-R category. Expansion of support provided two-parent families has long been on the frontier of welfare reform: one regular proposal, in part incorporated in the Family Support Act, is to mandate AFDC-U for all states. A more ambitious possibility is to eliminate the 100 hours rule. Because of the substantial AFDC-U samples created by several of these experiments, an opportunity was presented to learn more about these families and the consequences of employment assistance for their cohesion. To date, this opportunity has been bypassed and, as a consequence, the OBRA research results contributed little to the debate over AFDC-U extension that occurred in the formulation of the Family Support Act.

Impacts on Subgroups

Tables 9 and 10 provide impact estimates for selected subsets of treatment group members. In these tables, individuals are assigned to subgroups on the basis of their earnings during the year prior to becoming members of the research sample and on the basis of their welfare history prior to this point. Subgroups defined in this way are of interest because such criteria could be applied in the field. If individuals in a particular subgroup were found to be more responsive to the demonstration treatments than others, scarce program resources could be targeted at them.

In constructing Tables 9 and 10, we simply pieced together the published information available concerning subgroups, attempting to report estimates for as many OBRA demonstrations as possible. Most of the estimates that appear are from a 1988 MDRC report authored by Daniel Friedlander, although some are from MDRC evaluation reports on individual demonstrations. Because MDRC did not report subgroup estimates for all the demonstrations they evaluated, Tables 9 and 10 are not as complete as Tables 6 through 8. As can be seen, Tables 9 and 10 show treatment-comparison group differences in quarterly employment rates and present demonstration impacts on average amounts of AFDC received per calendar quarter. These two impact measures were selected because they are the ones most often used by MDRC in reporting program impacts on subgroups.

It is difficult to draw conclusions with any confidence concerning subgroup effects. Perhaps the most important observation that one can draw from Tables 9 and 10 is that no truly clear patterns concerning subgroup impacts emerge. There is some hint that, as compared to other subgroups, employment impacts are somewhat greater for applicants who have previously been on AFDC and somewhat smaller for applicants who have not. They are also, perhaps, smaller for applicants who earned over $3,000 during the year prior to application. There is also some suggestion that treatment impacts on AFDC payments were greater for treatment group members with modest amounts of prior earnings than for treatment group members with either relatively large amounts of prior earnings or no prior earnings at all. None of these relationships, however, is found consistently across all the demonstration programs. Moreover, even those patterns that perhaps can be discerned in Tables 9 and 10 must be interpreted cautiously. As pointed out earlier, both the number and types of services received by various subgroups may have differed. This may, in turn, have caused differences in subgroup impacts. Thus, it would appear that evidence from the OBRA demonstrations concerning targeting strategies is highly tentative at best.

An additional subgroup of interest, one not represented in Tables 9 and 10, is the heads of AFDC-R recipient units with children between three and five years of age. There has been concern that parents of young children could not be adequately served by WEPs because of difficulties in arranging for day care. As mentioned in Section II, prior to OBRA, such parents could not be required to participate in WIN. As Table 1 indicates, these parents were part of the client target pool in five of the OBRA demonstrations. The impact analyses for these demonstrations consistently implied that program effects on single parents with children between three and five were similar to those on single parents with only older children.

TABLE 9

Quarterly Employment Impacts on Subgroups from Selected MDRC Evaluations of OBRA Demonstrations

	Control Group: Average Percentage Employed in Absence of Program	Average Impact	Employment Impact of Programs on These Subgroups: Those with No Prior Earnings	Those with Prior Earnings $1 or More	Those with Prior Earnings $1 to $2999	Those with Prior Earnings $3000 or More	Those with No Prior AFDC Receipt	Those with Under 2 Years of Prior AFDC Receipt	Those with No Years or Under 2 Years of Prior AFDC Receipt	Those with Over 2 Years of Prior AFDC Receipt	Quarter(s) Over Which Impacts Averaged
AFDC-R applicants											
San Diego EPP/EWEP:											
Job search only	38.1%	(0.7)%	2.3%	(3.5)%	n.a.	n.a.	(5.0)%	5.3%	n.a.	(4.0)%	6 only
Job search/CWEP	37.4	4.5	7.5	n.a.	1.2%	1.9%	2.3	6.1	n.a.	4.8	4 - 6
Cook County:											
Job search only	32.4	(1.4)	(0.1)	(2.4)	n.a.	n.a.					4 - 6
Job search/CWEP	32.8	(1.1)	(1.5)	n.a.	(4.8)	1.7					4 - 6
Baltimore	42.2	4.3	4.1	n.a.	6.5	2.5	(0.4)	5.8	n.a.	5.6	4 - 10
Virginia	43.2	4.3	0.8	n.a.	4.2	9.5	1.9	7.0	n.a.	3.9	4 - 10
Arkansas	23.3	7.4	5.3	n.a.	13.6	1.7	4.9	10.2	n.a.	13.3	4 - 12
AFDC-R recipients											
Cook County:											
Job search only	17.3	1.6	0.5	6.3	n.a.	n.a.			n.a.		4 - 6
Job search/CWEP	17.9	1.9	1.3	4.2	n.a.	n.a.			n.a.		4 - 6
Baltimore	28.3	2.8	4.7	(2.4)	n.a.	n.a.		4.0	n.a.	2.4	4 - 10
Virginia	26.8	4.1	3.9	4.5	n.a.	n.a.		(0.2)	n.a.	5.8	4 - 10
Arkansas	10.5	2.5	3.4	(6.9)	n.a.	n.a.		2.7	n.a.	2.4	4 - 12
AFDC-R combined											
West Virginia	13.8	(0.4)	0.1	(2.8)	n.a.	n.a.	n.a.	n.a.	(0.2)%	(0.6)	6 only
AFDC-U applicants											
San Diego EPP/EWEP:											
Job search only	55.3	(1.5)	(0.9)	(1.7)	n.a.	n.a.	(3.4)	1.8	n.a.	(2.8)	6 only
Job search/CWEP	55.3	(2.2)	(6.3)	(0.5)	n.a.	n.a.	(4.7)	1.9	n.a.	(2.5)	6 only

Sources: Friedlander (1988), and various final reports on individual state work/welfare demonstrations published by the Manpower Demonstration Research Corporation.
Notes: See text for detailed explanation of table. Estimated effects reported in parentheses have unanticipated sign (that is, the treatment-control difference was negative).
n.a. = not available
Cells left blank where item not applicable.

TABLE 10

Impact on Quarterly AFDC Payments for Subgroups from Selected MDRC Evaluations of OBRA Demonstrations

	Control Group: Average Payments in Absence of Program	Average Impact	Payment Impact of Programs on These Subgroups: Those with No Prior Earnings	Those with Prior Earnings $1 or More	Those with Prior Earnings $1 to $2999	Those with Prior Earnings $3000 or More	Those with No Prior AFDC Receipt	Those with Under 2 Years of Prior AFDC Receipt	Those with No Years or Under 2 Years of Prior AFDC Receipt	Those with Over 2 Years of Prior AFDC Receipt	Quarter(s) Over Which Impacts Averaged
AFDC-R applicants											
San Diego EPP/EWEP:											
Job search only	$445	($2)	$25	($28)	n.a.	n.a.	($20)	$11	n.a.	$0	6 only
Job search/CWEP	469	33	40	n.a.	$63	($3)	5	74	n.a.	8	4 - 6
Cook County:											
Job search only	520	36	36	37	n.a.	n.a.					4 - 6
Job search/CWEP	485	15	21	n.a.	33	(3)					4 - 6
Baltimore	380	14	13	n.a.	38	(7)	9	8	n.a.	25	4 - 10
Virginia	210	19	26	n.a.	39	(8)	28	19	n.a.	14	4 - 10
Arkansas	156	26	22	n.a.	51	(1)	31	8	n.a.	75	4 - 12
AFDC-R recipients											
Cook County:											
Job search only	775	34	24	86	n.a.	n.a.					4 - 6
Job search/CWEP	744	13	6	40	n.a.	n.a.					4 - 6
Baltimore	622	(5)	(1)	(26)	n.a.	n.a.		25	n.a.	(19)	4 - 10
Virginia	436	24	26	16	n.a.	n.a.		(19)	n.a.	48	4 - 10
Arkansas	344	60	63	35	n.a.	n.a.		92	n.a.	44	4 - 12
AFDC-R combined											
West Virginia	341	16	17	10	n.a.	n.a.	n.a.	n.a.	$24	9	7
AFDC-U applicants											
San Diego EPP/EWEP:											
Job search only	471	85	95	81	n.a.	n.a.	24	146	n.a.	335	6 only
Job search/CWEP	471	65	160	28	n.a.	n.a.	37	108	n.a.	84	6 only

Sources: Friedlander (1988), and various final reports on individual state work/welfare demonstrations published by the Manpower Demonstration Research Corporation.
Notes: See text for detailed explanation of table. Estimated effects reported in parentheses have unanticipated sign (that is, the treatment-control difference was positive).
n.a. = not available
Cells left blank where item not applicable.

V. The Demonstrations: Evaluation

In a sense, a cost-benefit analysis of an OBRA demonstration can be viewed as a major end product of both the process and impact analyses that went before it. It provides a useful framework for organizing and summarizing what was learned in evaluating the impacts of the demonstration and for presenting these results to policymakers. Stated in the simplest terms, the objective of such an analysis is to measure all the costs and benefits associated with a demonstration program and determine whether the latter outweigh the former. In principle, if benefits are found to outweigh costs, the demonstration program should be made permanent. In practice, the objective of cost-benefit analyses of WEPs can never be fully realized; not all costs and benefits can be measured, and those that can are inevitably measured with at least some error. Thus, a cost-benefit analysis of an OBRA demonstration program cannot be used in a simple, straightforward manner to reach a go-no go decision concerning the program. Nevertheless, it can usefully facilitate that decision by organizing what is known and not known about the program, thereby permitting a rough assessment concerning the program's efficiency and focusing attention specifically upon those outcomes and costs that require political evaluation.

In this section, we assess the cost-benefit analyses that were conducted as part of the OBRA evaluations and summarize their major findings. Our discussion focuses mainly on those evaluations conducted by MDRC since, of all the OBRA evaluations, these incorporated the most comprehensive cost-benefit analyses by far. Indeed, many of the non-MDRC evaluations did not even include cost-benefit analyses, and those that did were typically limited to simple comparisons of direct budgetary expenditures on the treatment provided by a demonstration with the savings in AFDC and other welfare payments that treatment was believed to have generated. As will be seen, focusing only on welfare savings is overly narrow because it ignores other important benefits and costs associated with the demonstration programs.

Methodological Issues

The Cost-Benefit Framework

The basic cost-benefit accounting framework that was used by MDRC in its evaluations of the OBRA demonstrations was initially developed during the late 1970s and early 1980s for use in evaluating the Supported Work Demonstration. It has been used in most subsequent cost-benefit analyses of training, employment, and work programs directed at the poor. A stylized version of this framework appears in Table 11. Although details concerning the specifics of the cost-benefit typology MDRC employed vary somewhat from one WEP evaluation to another, depending upon the nature of the treatment and the vintage of the report, this table lists those benefit and cost components that were typically measured.

Table 11 presents benefits and costs from four separate perspectives: that of society as a whole; that of AFDC clients served by the evaluated demonstration program; that of nonparticipants, including taxpayers who paid for the program; and in terms of net effects on the combined budgets of local, state, and federal governments. Plus signs indicate anticipated sources of benefit and minus signs anticipated sources of costs from each vantage. Benefits and costs to society are simply the algebraic sum of benefits and costs to clients and to taxpayers, since both groups are part of society. Benefits and costs to clients are measured in terms of effects on their net money income. The table implies that if a demonstration program caused the AFDC payments of welfare recipients to decline, this should be regarded as a savings or benefit from a taxpayer and government budgetary perspective; a cost to program clients (albeit one that may be offset by earnings); and neither a benefit nor a cost from the perspective of society as a whole, but simply income transferred from one segment of the population to another. As the last two columns of the table imply, the taxpayer and budgetary perspectives are virtually identical, the only difference being the treatment of the value of any in-program output produced under demonstration programs--for example, by clients participating

TABLE 11

Stylized Cost-Benefit Framework for the OBRA Demonstrations

	Social (A) (B + C)	Client (B)	Taxpayer (C)	Budget (D)
Output produced by clients				
In-program output	+	0	+	0
Gross earnings	+	+	0	0
Fringe benefits	+	+	0	0
Client work-related expenditures				
Tax payments	0	–	+	+
Expenditures on child care, transportation, etc.	–	–	0	0
Use of transfer programs by clients				
AFDC payments	0	–	+	+
Other transfer payments	0	–	+	+
Program operating costs	+	0	+	+
Use of support programs by clients				
Support services received by clients	–	0	–	–
Allowances received by clients	0	+	–	–
Program operating costs	–	0	–	–

in CWEP. In their cost-benefit analyses, MDRC consistently reported findings from all four of these perspectives; most non-MDRC studies reported only budgetary consequences, if a cost-benefit analysis was included at all.

Table 11 divides the benefits and costs associated with the OBRA demonstrations into four major categories. The first two of these categories pertain to effects that result if an OBRA demonstration increases the work effort or productivity of participants--for example, by requiring them to work in a CWEP job where they perform useful services, providing them training, or helping them obtain private sector employment. On the one hand, the value of the output they produce will rise, which, in the private sector, should be reflected by increases in earnings and fringe benefits. On the other hand, if hours at work rise, expenditures on child care and transportation will also increase. And if earnings rise, tax payments will also increase. The third major cost-benefit category in Table 11 pertains to decreases in welfare dependency that may result from a WEP. Such reductions in dependency should cause both the amount of payments distributed under transfer programs and the cost of administering these programs to fall. The fourth major category refers to expenditures on support services for welfare recipients. Obviously, such expenditures increase when a demonstration program is implemented. However, this increase will be partially offset because participants in the demonstration programs no longer need to obtain similar services from previously existing programs.

Three of the subcategories listed in Table 11 are interrelated and require clarification: expenditures on child care, transportation, etc.; support services received by clients; and allowances received by clients. The first of these subcategories refers to total job-required outlays by clients on child care, transportation, uniforms, and other items. Support services pertains to the direct provision of such goods by a government agency, and allowances refers to government reimbursement of job-required expenditures by clients.

There appear to us to be two shortcomings in MDRC's treatment of these three subcategories. First, MDRC only measures job-required expenditures that occur while the client is participating in a demonstration program. In principle, job-required expenditures that occur after the client leaves the program should also be counted. Second, MDRC treats client job-required expenditures that are not reimbursed by the government as a social cost, but treats those expenditures that are reimbursed, as well as support services that are directly provided to clients, as transfers from taxpayers to clients that do not engender social costs. In our view, all job-required expenditures should be treated identically: as resource costs to society engendered in producing goods and services. Table 11 reflects this philosophy. These three benefit-cost subcategories just discussed involve relatively small amounts of dollars. Hence, alternative treatment of these three subcategories would not have had a major effect on findings from the cost-benefit studies conducted by MDRC.

In estimating each of the benefit and cost components that appear in Table 11, an issue arises that is very similar to one discussed in Section IV concerning the estimation of program impacts, namely, the unit of measure that should be used. There are at least four alternative possibilities: (1) dollars per AFDC recipient, (2) dollars per treatment group enrollee, (3) dollars per WEP participant, and (4) aggregate dollars (that is, dollars summed over all WEP participants). Since many AFDC recipients were ineligible for enrollment into the treatment group under the OBRA demonstration programs, and not all those who were enrolled actually received the treatments, benefit and cost estimates based on the first of the alternatives listed above will be smaller in magnitude than those based on the second, and those based on the second will be smaller than those based on the third. Benefits and costs based on the fourth alternative would, of course, be of a different order of magnitude than the other three.

In the OBRA evaluations it conducted, MDRC consistently reported its estimates of benefits and costs on a per-enrollee basis. Although this was certainly a reasonable choice, it is important to recognize that the magnitudes of MDRC's estimates of benefits and costs are influenced by this choice. Moreover, for reasons similar to those discussed in Section IV, the estimates are likely to be sensitive to program participation rates.

Benefits and costs that are sometimes referred to as "intangible effects," but are rarely, if ever, actually estimated in evaluations of WEPs, do not appear in Table 11. These include the values of leisure forgone and satisfaction gained from substitution of work for welfare. Even though techniques do exist for valuing forgone leisure (see Mishan, 1988, pp. 295-319 and Gramlich, 1981, pp. 72-75), intangible effects are almost by definition very difficult to measure. Although intangible effects are usually mentioned in MDRC's lists of caveats, they did not influence the ultimate bottom-line estimates of the total net gains or losses from the OBRA demonstration programs. Later, we examine the implications of not measuring these cost-benefit components.

A wide variety of data sources were used by MDRC in measuring the benefit and cost components listed in Table 11. For example, measures of program effects on earnings and transfer payment amounts relied on impact estimates of the sort described in Section IV, and estimates of program operating costs were based on administrative cost records. In some instances, special surveys were used. The procedures followed by MDRC in estimating each benefit and cost component are reasonable, and the work performed is solid and careful. Rather than detailing these procedures--an effort that would probably double the length of this manuscript--it is appropriate to refer the interested reader to the methodological synopsis in Long and Knox (1985).

We shall instead focus on several conceptual issues suggested by the cost-benefit accounting framework illustrated in Table 11. Before proceeding, however, it is important to emphasize that this framework offers several advantages: it is readily understandable to policymakers; by displaying benefits and costs from the perspectives of both clients and taxpayers, it suggests some of the distributional implications of the program being evaluated; and, possibly most important, since measures of each cost-benefit component listed in Table 11 can actually be obtained, it is operationally feasible. Indeed, as will be seen, it is far easier to find shortcomings in the framework than to suggest practical alternatives to it. However, the framework is not completely consistent with current cost-benefit theory, which typically describes benefits and costs in terms of compensating or equivalent variations or in terms of consumer or producer surpluses. Thus, it appears useful to compare some of the operational measures of benefits and costs used in OBRA evaluations with their conceptually correct counterparts.

This is accomplished in the next three subsections. In these subsections, most of the discussion focuses on benefits and costs associated with the OBRA demonstration component that came closest to the original workfare concept: mandatory community work experience programs. Each of the three subsections examines measures of costs and benefits associated with CWEP from a different perspective: the first subsection focuses on the client perspective; the second on the nonparticipant perspective; and the third on the social perspective.

<u>Benefits and Costs from a Client Perspective</u>

Figure 3, which is a standard labor-leisure indifference curve diagram, can be used to examine the major benefits and costs to clients from participating in programs such as those run under the OBRA demonstrations. This diagram is a slightly modified version of Figure 2, which was described in Section III. As the reader may recall, in Figure 3 the budget constraint that an AFDC client faces in the absence of welfare programs is represented by AM, and the budget constraint that she faces in the presence of welfare programs is represented by DFM. (For the moment, ignore budget constraint AN.)

Figure 3 is based on several simplifying assumptions. First, it is assumed that if the client represented in the diagram works at an unsubsidized job, she will earn the minimum wage. Second, it is assumed that the client is eligible for welfare, but that any earnings she receives are taxed at an

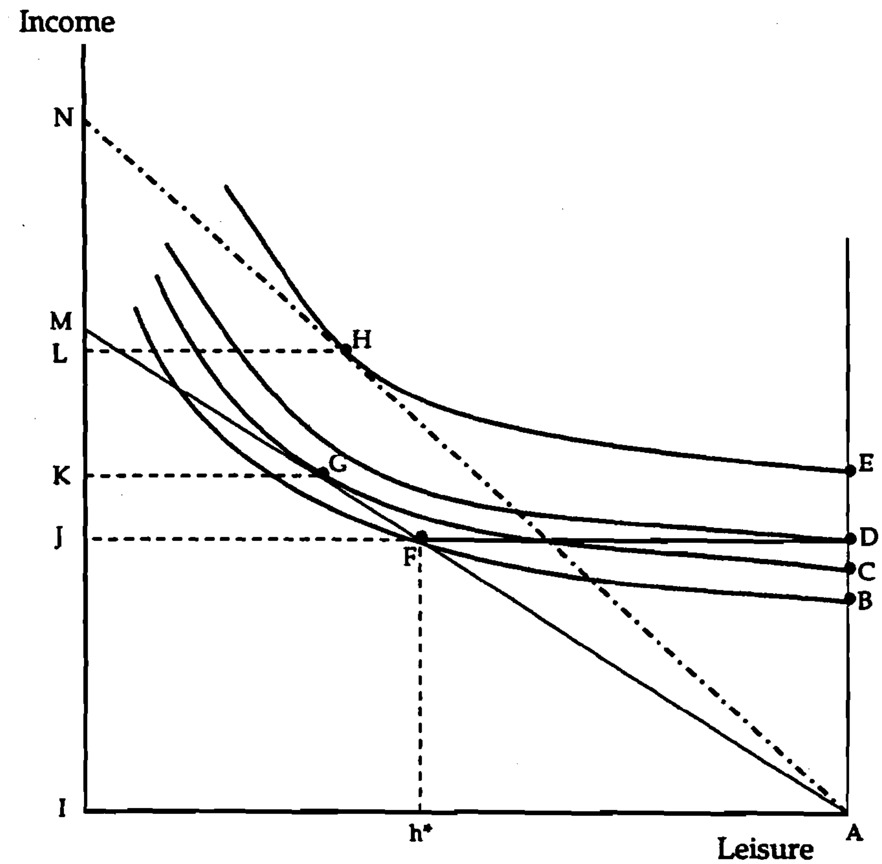

Figure 3: Client Behavior with Workfare and Training

implicit rate of 100 percent because every dollar of earnings reduces the welfare payment by one dollar. Third, it is assumed that, as a condition of receiving welfare, the client is required to work at a CWEP job for a number of hours, h*, that is determined by dividing her grant amount, JI, by the minimum wage. Finally, it is assumed that her indifference curves are negatively sloped throughout the relevant range. The last of these assumptions is a standard one. Later, however, we briefly deal with the alternative possibility: that utility may <u>increase</u> with work effort.

Although one could easily think of reasonable alternatives to the first three assumptions that appear above, together they fairly closely approximate the situation actually faced by many participants in those OBRA demonstrations that had mandatory CWEP components. In any event, the conclusions we will present below are insensitive to these assumptions; they are made purely for the sake of convenience, to keep Figure 3 from becoming overly cluttered and complex. For example, coupling the assumption that the client could potentially earn the minimum wage with the assumption concerning the CWEP hours requirement implies that the client would be located at the welfare program break-even point, F, while participating in CWEP. The assumption that the welfare program has a 100 percent implicit tax rate implies that the client would prefer point D--that is, zero hours of work--to either point F or point G, both of which are located on lower indifference curves.

The 100 percent tax rate assumption permits the equivalent variations (that is, changes in the client's welfare) associated with different changes in policy to be determined by comparing the points at which various indifference curves intersect with the right vertical axis of Figure 3. For example, the CWEP program, as we have seen, forces the client from point D to point F, which is located on a lower indifference curve. The client could be compensated for this loss and, hence, returned to her original indifference curve were she to receive an amount of money equal to DB. Thus, DB represents the equivalent variation associated with the imposition of the CWEP.

What is the implication of this for cost-benefit analyses of CWEP programs? Focusing on the client perspective, we answer this question by comparing the theoretically correct equivalent variation measure with the operational measure typically used in cost-benefit analyses of CWEP programs. This latter measure, which can be read off the left vertical axis of Figure 3, is simply the net change in client income that results from imposition of a CWEP program. As it happens, since neither transfer payments nor earnings change under CWEP, the client's net income, which in the diagram is measured as JI, also does not change. However, the diagram implies that from the perspective of the client, the CWEP program imposes a real cost equal to DB. Thus, according to the diagram, the cost to the client of working in a CWEP job will be understated by a conventional cost-benefit analyses of such programs.

Although the point just made--that standard cost-benefit procedures will measure the cost of CWEP to clients with error--is a valid one, it needs to be qualified in two ways. First, although program costs should, in theory, be measured in terms of either equivalent or compensating variation, both of these measures are, in fact, extremely difficult, if not impossible, to operationalize. Changes in income flows, in contrast, can be readily observed.

Second, the size, and even the direction, of the error that results from using observed changes in net income, rather than equivalent variation, as the cost-benefit measure depend on the assumption that time devoted to CWEP activities diminishes well-being--that is, the client's indifference curves have a negative slope. Assuming this is the case is consistent with the proposition that CWEP participation needs to be mandatory, at least for some. But for many welfare recipients meaningful work may in and of itself be a good. Indeed, MDRC surveys of CWEP participants (Friedlander et al., 1985; Friedlander et al., 1986; and Goldman et al., 1986) suggest considerable enthusiasm on the part of many for the opportunity. If this is the case, the preferences of clients would be better illustrated by indifference curves with positive, rather than negative, slopes over some interval to the left of the right axis in Figure 3. For such clients, participation in CWEP would result in a net gain, rather than impose a net loss. Note, however, that in either case, the standard cost-benefit procedure results in an error: client utility is changed by the program, but client income is not. The result is that standard cost-benefit procedure incorrectly assesses the outcome. Only if indifference curves are perfectly flat, so that the client is, in fact, completely indifferent between participating or not participating in CWEP, and no future personal gain is expected from the effort, will the operational cost-benefit measure correspond to the conceptually correct measure.

However, what may seem to be positive utility for devoting time to public service may instead reflect the perception on the part of CWEP participants that CWEP yields skills. Since the payoff to skills occurs in the future, such effects are difficult to capture in a single-period model. So far, the focus has been on benefits and costs to clients while they are participating in CWEP. Now, let us consider a different situation: the benefits and costs associated with a client move from welfare to an unsubsidized job. This is, of course, one of the major objectives of programs such as CWEP, and the OBRA demonstration findings described in Section IV suggest they are often successful in meeting this goal.

Figure 3 implies that CWEP would have its desired effect on the client represented in the diagram. Although in the absence of the CWEP program she would prefer not to work at all (point D), given a choice between participating in CWEP (point F) or working at an unsubsidized minimum wage job (point G), she would select the latter. In actual practice, the client might move directly from welfare to the job upon being confronted with the CWEP requirement or, alternatively, she might first participate in CWEP while seeking private sector employment.

A comparison of the operational cost-benefit measure of the client's move to private sector employment with the corresponding conceptually correct measure yields an interesting finding. On the one hand, the operational measure implies that the client enjoys a net gain. Transfer payments fall from JI to zero, but earnings increase from zero to KI, resulting in a net increase in income of KJ. On the other hand, measured in terms of equivalent variation, the client suffers a net loss equal to DC. Thus, under the particular circumstances represented in Figure 3, circumstances that seem typical of those facing many AFDC recipients, the operational measure suggests a conclusion that is the diametrical opposite from that implied by the conceptually correct measure. As before, the size and direction of the error depends upon the actual shape of the client's indifference curves. But only in the case of perfectly flat indifference curves would no error occur.

So far, operational and conceptual measures of benefits and costs have been compared under two sets of circumstances: while an AFDC client is actually participating in CWEP; and when CWEP causes the client to accept a job similar to one she could have obtained, but would have elected not to take, in the absence of the program. There is a third possibility. CWEP, perhaps, augmented by job search or training, might help the client secure a better job than would have been available in the absence of the program. In terms of Figure 3, the post-CWEP period in this case is represented as a move from budget constraint AM to budget constraint AN. This change in the budget constraint facing the client permits the client to reach equilibrium at point H, rather than at point D. Given these circumstances, the client is clearly better off. As can be seen, the equivalent variation associated with the program, which is equal to ED, is positive. However, the client's net gain in income, LJ, is even larger. Hence, as we found in the previous two cases, as long as the client's indifference curves are negatively sloped, the operational measure of the client's benefits and costs will overstate her true net gain (or understate her true net loss).

In the case in which CWEP holds promise of new skills or opportunities, the behavioral consequences of these effects in a single period might best be simulated by drawing line DJ in Figure 3 with a positive slope, reflecting the increased real current "income" generated by the changed future prospects the WEP brings about. This "shadow wage" is the present value of compensated future CWEP-generated earnings gains that result from one more unit of CWEP effort.

Benefits and Costs from the Nonparticipant Perspective

1. Tangible Benefits Received by Nonparticipants. The preceding section emphasized that in measuring the benefits and costs of WEPs to clients, it is changes in utility that should be estimated, rather than changes in money income. Thus, in principle, the effects on client utility of WEP-induced reductions in the leisure time of program participants should be taken into account.

The same concept applies in measuring benefits and costs of WEPs to nonparticipants. In particular, it seems possible that WEP-induced reductions in the leisure time of welfare recipients may increase the utility of some nonparticipants, especially those nonparticipants who pay the taxes used to support the welfare system. This is an intangible benefit of WEPs to nonparticipants that should, in principle, be taken into account in conducting cost-benefit analyses of these programs, although given the practical difficulty of measuring the magnitude of this benefit, it has never been done. Nevertheless, it is important to recognize that WEP-induced reductions in client leisure time may have positive effects on the utility of nonparticipants that to some unknown extent offset negative effects on client utility.

2. In-Program Output. This subsection focuses on measuring the value to taxpayers of the in-program output produced by participants in CWEP. Ideally, this would have been done by determining what taxpayers would be willing to pay for this output. This was infeasible, however, since the output was not purchased in market transactions. Consequently, MDRC used an alternative approach. This approach involved attempting to determine what the labor resources required to produce the output would have cost if purchased on the open market. This is consistent with the procedures used for measurement of public sector output in the national income accounts. However, in this case, evaluation of output on the basis of resource cost is complicated by the fact that the government and nonprofit agencies that "employed" CWEP participants paid nothing for the services of these people. Thus, the wage rate that would have been paid similar workers hired in the open market to do the work performed by the CWEP participants had to be determined. Since most CWEP participants worked in entry-level jobs, the wage rate used for this purpose typically was not far above the minimum.

Once an appropriate wage rate was determined, the basic calculation involved multiplying the number of hours AFDC recipients worked in CWEP jobs by this hourly wage, adding normal fringe benefits, and inflating the total by the ratio of the average productivity of CWEP workers to the average productivity of similar regular workers. Interestingly, this ratio, which was determined by work-site surveys administered to the supervisors of CWEP participants, generally slightly exceeded one, indicating that, on average, CWEP workers were considered a bit more productive than workers hired through regular channels.

The procedure just described can result in an estimate that either overstates or understates the true value of the in-program output produced by CWEP participants. The reasons for this can be seen by examining a key assumption that implicitly underlies the valuing method that MDRC uses: that the decisions of the government and nonprofit agencies that employ CWEP workers closely reflect the desires of taxpayers. More specifically, an analogy is implicitly drawn with the behavior of private sector firms and consumers under perfect competition, and it is assumed that the amount that an agency would be willing to pay to employ an additional worker corresponds to the value that taxpayers would place on the additional output that the worker could potentially produce. Although this is not an appropriate place to assess the perfect competition analogy or discuss the extent to which bureaucratic behavior reflects taxpayer preferences, it should be obvious that this assumption is a rather strong one.

The implications of the assumption can be explored by use of Figure 4, which depicts the demand curve for low-skilled workers of a government or nonprofit agency that might potentially be assigned CWEP workers and the supply curve the agency faces in hiring low-skilled workers in a competitive labor market. In using this diagram, let us first examine a situation where the bureaucratic behavior assumption is valid and then a case where it is not.

In Figure 4, the horizontal line, S, represents the supply curve, which is set at the level of the market-determined price (that is, wages plus fringes) that must be paid each worker; and the downward sloping line, D, represents the demand curve, which is assumed to slope downward as a

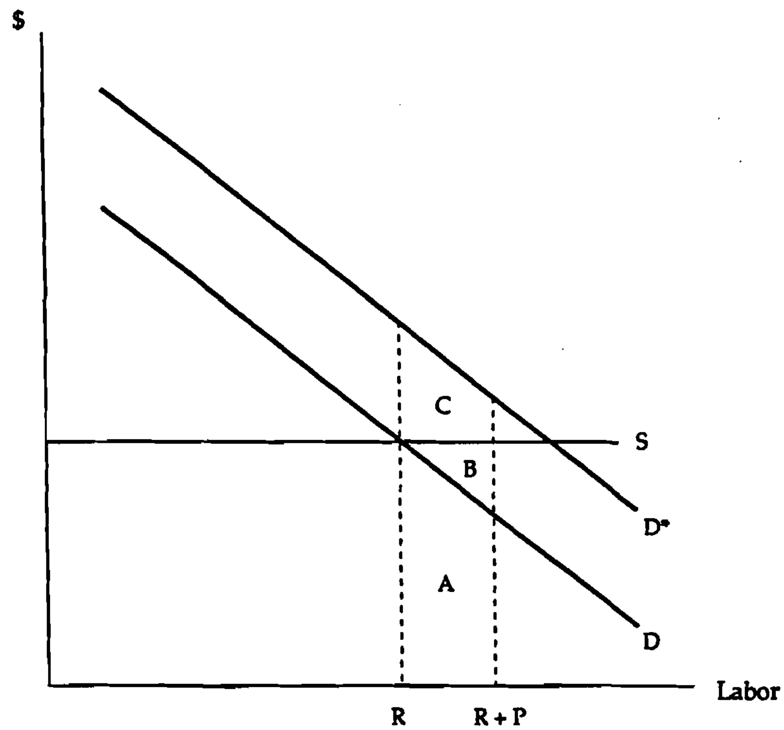

Figure 4: Labor Demand and Supply

result of diminishing returns and (as implied by the bureaucratic behavior assumption) because the agency prioritizes its tasks so that, as its budget expands, successively less-important services are performed. (Ignore curve D* for the moment.) This demand curve reflects the willingness to pay for workers by the agency, and, in keeping with the assumption concerning bureaucratic behavior, the area under this curve is presumed to measure the value to taxpayers of output produced by workers hired by the agency.

Figure 4 indicates that, in the absence of CWEP workers, the agency would hire R regular workers; but if P CWEP participants were assigned to the agency, a total of R+P workers would be employed. Thus, if the bureaucratic behavior assumption is valid, the value to taxpayers of the output added by the CWEP workers would equal area A.

Unfortunately, however, area A typically cannot be directly measured. The reason for this is that the output produced by government and nonprofit agencies is rarely sold in market transactions, and, consequently, the demand curve depicted in Figure 4 cannot be observed. However, even though government and nonprofit agencies that "employ" CWEP participants pay nothing for the services of these people, the area under the supply curve between R and R+P can be valued by simply determining the wages and fringe benefits that would have to be paid to similar workers hired on the open market to do the work performed by CWEP participants. Consequently, it is the area under the supply curve--that is, area A <u>plus</u> area B--that is usually actually used as the measure of the value of worker output in CWEP. As a glance at Figure 4 suggests, the size of the resulting overstatement of the value of the output produced by CWEP participation, which is represented by area B, depends upon the slope of the demand curve.

So far, we have assumed that agency behavior simply reflects the value that taxpayers place on the agency's output. Let us now look at one of the numerous possible situations where this need not be the case. The specific example we examine is one in which the agency, perhaps because of budget constraints resulting from the public goods characteristics of whatever services it provides, produces less output than taxpayers collectively desire. These circumstances are represented in Figure 4 by two demand curves. As before, curve D indicates <u>agency</u> willingness to pay for workers, but the value that <u>taxpayers</u> place on the output produced by the agency is now represented by the area under demand curve D*. Consequently, the value of the additional output produced by the P CWEP participants provided the agency now equals area A plus area B plus area C. Thus, under these circumstances, MDRC's measure of the output produced by CWEP workers, which as previously indicated equals area A plus area B, understates the true value. The size of this understatement is equal to area C.

3. Public Sector Displacement. To this point, our discussion has been based on the assumption that the CWEP workers made available to an agency would simply be added to the regular work force that the agency would hire in the absence of CWEP. This need not be the case. The agency might instead substitute CWEP workers for regular workers. In terms of Figure 4, this behavior on the part of the agency, which is usually referred to as "displacement," would mean that agency employment would increase by less than the P number of workers provided by CWEP. Indeed, with 100 percent displacement, the agency's work force would remain at R, rather than increase to R + P. Consequently, displacement leads to overstatement of the value of output produced by CWEP participants. When displacement occurs, the money saved by local government is used either for reducing taxes, reducing the net deficit (increasing the net surplus) of government, or enhancing other types of services. In the tax or surplus case, the logic of cost-benefit analysis treats the transaction as only a transfer, and the only impact of displacement on CWEP evaluation is to reduce benefits accruing to taxpayers. When savings generated by displacement are used for purchase of other services, those services themselves are a benefit.

The distortion in benefit assessment caused by displacement obviously depends upon the amount of displacement that actually occurs. This issue was investigated in several of the OBRA evaluations through surveys of the agencies employing CWEP workers, and displacement was consistently found to be very small. Perhaps the most convincing examination of this issue was conducted as part of the evaluation of North Carolina's Community Work Experience demonstration project (North Carolina Department of Human Resources, 1985, pp. 19-20), where it was found that the number of positions for regular workers at agencies employing CWEP workers actually increased during the demonstration period and that job vacancy rates for regular workers remained very low. It should be emphasized, however, that while these findings are certainly suggestive, they are far from definitive. For example, perhaps the number of slots for regular employees would have expanded by an even greater amount in the absence of the demonstration program.[8] Moreover, like all other examinations of displacement, the North Carolina findings are limited to the demonstration period. Long-run displacement under a permanent program could be far greater than that which occurred during a two-year demonstration.

While displacement is often a highly charged political issue, its importance may be exaggerated. Debate over displacement reflects three issues. The earliest is a carryover from the era of use of public sector employment as a macroeconomic countercyclical demand-management tool. It is sometimes claimed that stimulation of public sector employment has advantages as a means of increasing aggregate demand in the environment of a general business recession (Wiseman, 1976). If displacement occurs, the associated fiscal stimulus is lost. We believe it is fair to say that recent development in macroeconomic theory plus empirical study of the operation of countercyclical public employment policy have seriously weakened, if not destroyed, the credibility of the arguments for such policies. In any event, economic stabilization is not a cornerstone of the case for employment-oriented welfare reform.

The second and third issues both involve the nature of the change in public sector employment demand brought about by CWEP-type programs. There is some evidence that employees in public sector jobs tend to receive wages in excess of those that persons with comparable skills receive in private employment. If this is the case, displacement reduces the number of such jobs and the beneficiaries of them: existing or potential public sector employees lose income as a result but, of course, taxpayers as a class gain.

Beyond this wage effect, displacement presumably substitutes a welfare recipient for someone who does not receive welfare. If the displaced worker has a similar risk of joblessness, increases in output produced by welfare recipients assigned to CWEP may be offset by losses in the output formerly produced by the workers displaced by CWEP. This need not be the case, however, if

CWEP participants are exceptional in terms of lack of skills or if they face exceptional barriers to labor market access. Moving such persons directly into jobs under these circumstances will change the characteristics of the general pool of unemployed in ways that may allow a reduction in the net incidence of joblessness. Thus, to the extent that CWEP clients are unskilled relative to those they replaced, the labor market consequences will differ from a simple one-for-one replacement.

Technology probably also retards net displacement. CWEP employment features high turnover by design. High-turnover jobs require greater management, thereby increasing jobs for employees in the next tier up.

4. Private Sector Displacement. A major objective of most of the OBRA WEP demonstrations was to increase the unsubsidized private sector employment of program participants. To the extent these efforts were successful, some participants undoubtedly ended up in jobs that would otherwise have been held by nonparticipants. If, as a result, these nonparticipants became unemployed or accepted lower-wage jobs, their earnings obviously fell. This earnings reduction, which is another type of displacement effect, was potentially a cost of the OBRA demonstrations to nonparticipants or taxpayers, a cost that typically has not been measured in cost-benefit analyses of these programs.

Here, again, the net effect on employment depends, in part, upon the consequences for the local labor market of the change in the composition of the pool of unemployed that is brought about by each WEP demonstration. It also depends upon local labor market conditions during the demonstration period. For example, if local unemployment was low, it should have been relatively easy for displaced nonparticipants to find alternative job opportunities. Consequently, any displacement effects should have been small. But if unemployment was high and local labor market conditions were loose, which, in fact, was the situation during at least the earlier phases of many of the OBRA demonstrations, displacement effects could have been substantial.

Gramlich (1981, pp. 61-67), has argued that any displacement effects could, in principle, be eliminated by appropriate use of monetary and fiscal policy. He goes on to suggest that the only real constraint on using macroeconomic policies to do this are the inflationary pressures that might be engendered. The implication of this is that if inflationary pressures were sufficiently small at the time the OBRA demonstrations were conducted, which was probably generally the case, it would have been feasible to use macroeconomic policy to offset any private sector displacement resulting from the demonstrations. Thus, Gramlich provides a theoretical argument that would appear to justify ignoring displacement effects in cost-benefit analyses of the OBRA demonstrations. This is reassuring because, as noted above, these effects were, in fact, ignored.

Benefits and Costs from the Social Perspective

Since the social perspective provides the most comprehensive framework from which to assess program effects and is often used to determine whether a program increases or decreases overall economic efficiency, the cost-benefit literature generally tends to emphasize this view. In its reports, however, MDRC appears to take a somewhat agnostic stance concerning the most relevant perspective. In our opinion, given certain problems that are inherent in the computation of benefits and costs from the social perspective, this stance is not unreasonable. In this subsection, we briefly discuss these problems.

One issue concerns the fact that by focusing on clients and taxpayers, cost-benefit analyses of the OBRA demonstrations have tended to ignore program effects on the welfare of a third potentially important group: the children of clients. For example, by increasing the time that clients spend outside the household, the OBRA treatments reduced parental attention received by their children. To the extent this effect was deleterious, it was obviously most likely to be so in one-parent households with very young children. On the other hand, by increasing client attachment to the work force, the treatment programs may have helped parents become better role models for their children. Neither of these effects are incorporated into the cost-benefit analyses. The reasons for this are apparent: (1) we do not know what the effects on children are; and (2) if we knew, they would be difficult to value.

A second issue concerns the method that MDRC used in computing social benefits and costs. As previously indicated, this was simply done by algebraically summing gains and losses to clients with those to nonparticipants. In this approach, a dollar gained or lost by a client is treated as a dollar gained or lost by a nonparticipant. Consequently, for example, if a WEP caused the transfer dollars received by AFDC recipients to fall, this would be viewed as not affecting society as a whole, since the loss to clients would be fully offset by benefits to nonparticipants in the form of reductions in government budgetary outlays. As is true of most cost-benefit analyses, if the WEP is found to produce a positive net gain for society as a whole, this can be interpreted as indicating that total available social output has increased and, consequently, that the program has resulted in an improvement in overall economic efficiency.

A shortcoming inherent in this approach can be seen by use of a stark example. Suppose a state is considering completely abolishing its AFDC program and conducts a cost-benefit analysis to help assess the potential consequences of such a policy. Using the standard approach, the state would probably find that, from a social perspective, the benefits from eliminating AFDC would far exceed the costs. This conclusion is implied by the fact that the very existence of transfer programs such as AFDC cause economic inefficiency--resources must be expended in operating these programs and labor supply is reduced by them.

The conclusion is also suggested by an examination of the cost-benefit components listed in Table 11, which indicates that eliminating AFDC engenders two social benefits and two social costs. The two social benefits are the elimination of AFDC operating expenditures and the rise in earnings that would occur as many former AFDC recipients are forced to increase the number of hours they work. These social benefits would be offset, but probably only in small part, by expenditure increases on child care and transportation and by some increase in the use of social service programs by former recipients.

If our assertion that a traditional cost-benefit study would indicate that society would be better off if the AFDC program were completely eliminated is correct, a natural question to ask is why this program continues to exist. One possible answer is that society, at least as its view is expressed through the political process, would not accept the conclusion implied by such a cost-benefit study. This, in turn, suggests that the convention upon which the social perspective calculations are based--that a dollar gained or lost by an AFDC recipient is equivalent to a dollar gained or lost by a nonparticipant--does not accurately reflect actual social preferences. The very existence of a program such as AFDC suggests that society is willing to sacrifice efficiency to increase the incomes of the low-income persons who qualify for program benefits.

In the cost-benefit literature, an issue such as the one just raised is typically treated through the use of distributional weights that, in principle, reflect society's view as to what constitutes an equitable distribution of income. In the context of the OBRA demonstrations, for example, the gains and losses of clients would be given a weight greater than one and those of nonparticipants a weight less than one. Once this weighting was completed, benefits and costs from the social perspective could then be computed. The obvious operational problem with doing this, however, is that the values of the weights appropriate for this purpose are unknown. This, however, does not justify the standard approach in which clients and nonparticipants are treated identically and, consequently, the gains and losses of both groups are implicitly given a weight equal to one.

There are at least two reasonable ways in which the lack of information concerning appropriate weights might be handled. The first is to select several of the alternative weighting schemes suggested in the literature (see, for example, Gramlich, 1981, pp. 118-123 and Pearce, 1983, pp. 59-71) and then use these to conduct sensitivity tests. A second approach is simply to report unweighted benefits and costs for program participants and nonparticipants, but not for society as a whole--thereby requiring policymakers to apply their own set of implicit weights in interpreting these results. This would seem especially appropriate if one acknowledges that the public interest in WEPs arises not only from fiscal concerns, but from qualitative aspects of the circumstances of welfare recipients as well. Even if a WEP is not cost-effective as measured by the standard procedure, it might be preferred over the absence of intervention if, for example, it resulted in a reduction in welfare dependency.

The pretention of commensurability is the great advantage of cost-benefit analysis. By raising the possibility that in fact a common unit of exchange cannot be found for all benefits and costs, we are faced with the question of what should and should not be reported. Here, we believe the MDRC example is sensible: in addition to its version of cost-benefit analysis, they present the net effects of each program on client experience within the system (process) and the quantitative consequences for employment rates and the incidence of welfare receipt. These seem to be the matters that generate most concern in welfare policy.

Projections of Benefits and Costs

In conducting cost-benefit analyses of the OBRA demonstrations, it was necessary to take account of the fact that some benefits and costs extended beyond the demonstration period. For example, as a result of having participated in an OBRA demonstration, some individuals could potentially enjoy increased earnings, but pay higher taxes, incur greater job-required expenses, and receive fewer transfer payments over the remainder of their working lives. These streams of future benefits and costs must be incorporated into the cost-benefit analyses. Doing this requires that three important parameters be specified: the time horizon, the discount rate, and the decay rate. Each of these parameters will be discussed in turn.

The time horizon is the time period over which benefit and cost streams are estimated. The standard procedure for determining the length of this period in evaluations of human resource programs is to subtract the age of program participants at the time they entered the program from the age at which they are expected to retire from the work force. In its cost-benefit analyses of the OBRA demonstrations, however, MDRC used a shorter, somewhat arbitrarily selected time horizon of five years. Since MDRC typically obtained follow-up information on program participants for an "observation period" of around two years, this meant that it was necessary to project benefits and costs for only around three years.

The use of a five-year time horizon was simply an acknowledgement on MDRC's part that since cost-benefit analysts do not possess crystal balls, uncertainty increases the further one attempts to extrapolate beyond the observation period. Basing a cost-benefit analysis on a short time-horizon, however, can potentially distort its findings. The reason for this is that some benefits and costs (for example, program operating costs) do not have to be extrapolated at all, while others (earnings improvements) may actually persist well beyond five years. Use of a short time-horizon tends to give less weight than appropriate to benefits and costs that remain important for long periods of time.

The second parameter that is needed in projecting benefit and cost streams is a discount rate. Use of a discount rate is required to compare dollars received or paid in future years with those received or paid during the demonstration period. Choice of discount rate in cost-benefit analysis is controversial. In actual practice, however, cost-benefit analysts have rarely used discount rates that are smaller than 3 percent or that exceed 10 percent, and analyses of human resource programs are, perhaps, most often based on values of between 4 and 6 percent. In its OBRA evaluations, MDRC uses a value of 5 percent.

A third parameter, a decay rate, is necessary in extrapolating demonstration effects that persist beyond the observation period in order to take account of the possibility that the size of these effects may vary over time. It is often argued, for example, that programs that provide training or job placement for low-wage workers initially give those who received these services a competitive advantage in the labor market, but that this advantage "decays" over time. In the case of training, however, one could alternatively argue that doors are opened on the job that allow participants to obtain additional training even after leaving a program, and, consequently, the program's effects on earnings will grow over time.

Unfortunately, only very limited empirical evidence exists as to whether the earnings effects of treatments such as those offered by the OBRA demonstrations tend to grow or decay over time, let alone what the actual rate of decay or growth is. Thus, the choice of an appropriate decay rate was one of the more problematic aspects of the OBRA cost-benefit analyses.

In the analyses it conducted, MDRC generally selected several alternative values for the decay rate and subjected its findings to sensitivity tests. The exact values used for this purpose varied among the studies but usually included a decay rate of zero, which implied no decay, and often also a rate of infinity, which implied that no benefits or costs extended beyond the observation period. In addition, the actual rate of decay that occurred during the demonstration observation period was sometimes used, as was a rate of 22 percent, which was obtained from a 1980 national study of the WIN program conducted by Ketron (1980). In no case did MDRC use a negative decay rate--that is, a rate that implied that benefits or costs grow, rather than decline, over time. However, as indicated by Table 6 in Section IV, there is some evidence from MDRC's own data that such growth did, in fact, occur in some of the OBRA demonstrations.

Table 12 presents results from a sensitivity test MDRC conducted on cost-benefit estimates derived for the Baltimore Options program. These particular results were selected because they could be derived from two separate sets of published estimates: one that appeared in MDRC's 1985 final report for the Baltimore Options program (Friedlander et al., 1985); and one from a 1987 supplemental report that updated the original estimates by extending the observation period upon which they were based and, hence, shortened the length of time over which extrapolation was required (Friedlander, 1987). This is the only instance with which we are familiar in which MDRC has updated its initially published cost-benefit estimates. By comparing the two sets of estimates, one can obtain some sense of how sensitive the cost-benefit results are to incorrect assumptions concerning the rate of decay.

The estimates appearing in Table 12 pertain to the client perspective and are averaged over all members of the AFDC-R treatment group. As can be seen, three pairs of columns are displayed, one for each alternative assumption concerning decay rates. The left column in each pair is taken from MDRC's originally published estimates and the right column from the updated estimates.

In Table 12, some variation is apparent across the three pairs of columns, but even greater differences occur within each pair. These differences between the original and updated estimates are attributable to the fact that at least during the extended observation period, even an assumption of a zero decay rate for the earnings impact of the Baltimore Options program was overly pessimistic. As it turned out, the earnings differences between the treatment and comparison groups in Baltimore continued to grow during the extended observation period.

Given the absence of good evidence concerning decay rates, MDRC's strategy of conducting sensitivity tests with alternative rates was a good one. It is important to recognize, however, that the amount of variance observed in these sensitivity tests was severely restricted by the very short five-year time horizon upon which they were all based. Nevertheless, as Table 12 attests, this variation was far from eliminated.

Findings from the Cost-Benefit Analyses

Table 13, which is somewhat similar in format to Tables 6 and 7, presents summary statistics from the cost-benefit analyses of the OBRA demonstrations conducted by MDRC. As in Tables 6 and

TABLE 12

Six Alternative Estimates of Benefits and Costs per Baltimore Options Program Treatment Group Member: Reported from the Client Perspective

	Lower Estimates		Middle Estimates		Upper Estimates	
	Original Estimates	Updated Estimates	Original Estimates	Updated Estimates	Original Estimates	Updated Estimates
Earnings and fringe benefits	$491	$1,277	$930	$1,886	$1,272	$2,021
Tax payments	-81	-172	-247	-269	-343	-290
Expenditures on child care and transportation	-21	-25	-24	-26	-24	-26
AFDC payments	-29	-52	-100	1	-148	13
Other transfer payments	-111	-174	-225	-211	-297	-213
Allowances and support services	190	220	213	234	213	234
Net gain	439	1,074	547	1,615	673	1,739

Source: Original estimates are from Table 6.7 of the Maryland Final Report. Updated estimates are from Table A.7 of the Maryland Supplemental Report.

Notes: Lower estimates represent only observed program impacts, and thus do not include estimates of future impacts. That is, there is no extrapolation beyond the observation period. Middle estimates include estimates of future benefits and costs that are based on an assumed annual decay rate of 22 percent. Upper estimates include estimates of future benefits and costs that are based on an assumed zero decay rate. Original estimates are based on an observation period of around one and one-half years and an extrapolation period of about three and one-half years. Updated estimates are based on an observation period of around three years and an extrapolation period of about two years.

7, MDRC's estimates are reported separately for AFDC-R and AFDC-U households. For those demonstrations for which MDRC provided the necessary information, results are also reported separately for new AFDC applicants and prior AFDC recipients. Parentheses are used to indicate when benefits or costs are in the opposite direction from that intended by the OBRA demonstration programs--for example, when net remuneration from employment declines or amounts of transfer payments increase.

Two sets of columns appear in Table 13. The first presents estimated benefits and costs from the client perspective and the second from the nonparticipant perspective. The first column in each set reports total net gains (or losses) from each of the two perspectives, while the remaining columns provide information on the benefit and cost components that together account for these gains (or losses). For example, column B reports the estimated net gain by clients from employment under each demonstration program--that is, estimates of the sum of any increases in earnings, fringe benefits, and any work-related allowances paid under the program less the sum of tax payments and client job-required expenditures on such things as child care and transportation. Column C indicates changes in the amount of AFDC and other transfer benefits received by clients. Column E presents MDRC's valuations of output produced under CWEP. Column F is the sum of tax increases paid by clients, reductions in transfer payments paid to clients, and reductions in transfer program operating costs, all of which may be viewed as benefits from the nonparticipant perspective. Finally, column G shows the cost of the treatment programs to nonparticipant taxpayers. This column, which is computed by summing program operating costs and the cost of support services and allowances provided clients, is identical to the last column in Table 8, except for the New Jersey and Maine demonstrations. For these two, Table 8 includes wage subsidies paid to employers, while Table 13 excludes these subsidy amounts. The reason for this difference is that although subsidies paid to employers are a cost to the government of running a treatment program, from a nonparticipant

TABLE 13

Summary of Cost-Benefit Estimates from MDRC Evaluations of OBRA Demonstrations

	Client Perspective (Cost-Benefit Component)			Nonparticipant Perspective (Cost-Benefit Component)			
	Net Present Value (A=B-C)	Net Gains from Employment	Loss in Transfer Payments	Net Present Value (D=E+F-G)	Value of CWEP Output	Tax-Transfer Gains	Service Program Costs
	A	B	C	D	E	F	G
AFDC-R							
Applicants							
San Diego EPP/EWEP:							
Job search only	$ 644	$1,323	$ 679	$ 452	($3)	$ 965	$ 510
Job search/CWEP	798	1,874	1,076	1,156	205	1,529	578
San Diego SWIM	(880)	1,140	2,020	1,633	180[a]	2,153	700
Virginia	1,134	1,698	564	667	41	879	253
West Virginia	(481)	(407)	74	389	542	25	178
Recipients							
San Diego SWIM	725	3,158	2,433	1,698	180[a]	2,586	1,068
Virginia	574	982	408	190	145	593	548
West Virginia	80	157	77	873	1,059	115	301
New Jersey	1,262	2,278	1,016	1,069	(9)	1,591	513
Maine	3,182	4,497	1,315	(418)	680	894	1,992
Applicants & recipients							
Cook County:							
Job search only	(420)	145	565	475	1	601	127
Job search/CWEP	(34)	311	345	362	100	420	158
Baltimore	1,739	1,939	200	74	390	513	829
Arkansas	(449)	410	859	944	20	1,082	158
AFDC-U							
Applicants							
San Diego EPP/EWEP:							
Job search only	(1,196)	375	1,571	1,229	(5)	1,777	543
Job search/CWEP	(1,443)	129	1,572	1,414	354	1,732	672
San Diego SWIM	543	2,083	1,540	1,577	267[a]	1,970	660
Recipients							
San Diego SWIM	(921)	2,178	3,099	2,487	267[a]	3,242	1,025
Applicants & recipients							
Baltimore	(1,223)	(2,017)	(784)	(1,856)	280	(1,599)	537

Source: Various final and supplemental reports on individual OBRA demonstrations published by the Manpower Demonstration Research Corporation.

Notes: See text for detailed explanation of table.

[a]Separate estimates of the value of CWEP output for applicants and recipients not provided for San Diego.

( ): Negative net gains or increases in transfer payment amounts.

perspective they are best viewed as a transfer from one group of nonparticipants to another.

The net gains and losses implied by Table 13 are not especially large. For example, a program that resulted in a net client or nonparticipant gain of $1,000 per treatment group member (a figure that is larger than most of those appearing in either column A or D) and enrolled one million AFDC recipients each year (a figure that seems improbably large) would produce a total annual gain of $1 billion, a gain that may be usefully compared to the budgetary cost of AFDC, which is currently around $16 billion per year (Burtless, 1990, Table 1).

Perhaps the most striking finding from Table 13 is that of the nineteen reported sets of cost-benefit estimates, seventeen imply net gains for nonparticipants, but nine indicate that clients suffered net losses. The reason for the net losses among clients, which occur disproportionately among the AFDC-U population, is suggested by a comparison of columns B and C. It appears that some of the demonstration treatments did not result in sufficiently large gains in employment to offset client losses in transfer payment receipts. Thus, Table 13 implies that in eight of the nine instances nonparticipants gained at the expense of clients. In one case, Maine, clients actually gained while nonparticipants suffered losses. Nine of the ten remaining sets of cost-benefit estimates imply that the demonstration treatment made both clients and nonparticipants better off; the tenth, that for Baltimore's AFDC-U population, indicates that both groups were made worse off.

The reader may notice that Table 13 does not present estimates of net gains and losses from the social perspective. The reader is invited to do this for himself or herself by summing columns A and D, using whatever distributional weights he or she feels are appropriate. The choice of distributional weights will not affect the ultimate conclusion that one draws in the ten instances in which net treatment effects on clients and nonparticipants are in the same direction, but may in at least some of the nine cases in which they are in the opposite direction.

The preceding two paragraphs interpret the benefit and cost estimates appearing in Table 13 strictly at face value. Any final assessment of these results must take account of the difficulties inherent in measuring benefits and costs associated with the OBRA demonstrations. In particular, it was suggested earlier in this section that if indifference curves representing the trade-off between income and leisure are negatively sloped, which is the standard assumption, net gains for clients would be overstated (or, alternatively, net losses understated). Thus, some of the ten instances of measured net gains for clients that appear in Table 13 could actually be net losses. On the other hand, if reductions in client leisure increase the utility of nonparticipants, Table 13 would understate the net gains enjoyed by these persons. It was also suggested earlier that the estimate of CWEP output values could be either overstated or understated, depending on whether the agencies employing CWEP participants produce the amount of output desired by nonparticipants and whether workers hired through normal channels by these agencies are displaced by CWEP participants. It should be noted, however, that except for the demonstrations conducted in West Virginia, Maine, and Baltimore, even a halving or a doubling of the CWEP output values presented in column E of Table 13 would have only modest impacts on the sizes of the total net gain estimates appearing in column D.

Conclusions

This paper has suggested that cost-benefit analyses of WEPs are more difficult to both conduct successfully and interpret than they may first appear. Nevertheless, they can provide useful insights. For example, a major impetus for many of the WEP demonstrations has been to reduce the size of welfare caseloads, and there is considerable evidence from impact analyses that they have been modestly successful in accomplishing this. However, cost-benefit analysis suggests that the net social gains from doing this are generally very small, and if appropriate distributional weights are used, may often be negative. Moreover, analysis from the client perspective implies that WEPs make many welfare recipients worse off, especially when the value of their leisure time is taken into account.

VI. Reflections on the Outcomes

We return now to the question introduced at the beginning: why was welfare reform successfully accomplished in the 1980s, under fiscal circumstances far less hospitable than those which attended the failure of the Family Assistance Plan and the Program for Better Jobs and Income? More to the point, what, if any, was the role of the OBRA demonstrations in bringing reform about?

OBRA and the Family Support Act

At first review, the Family Support Act (FSA) seems to bear only a tenuous relation to the state initiatives we have reviewed. The legislation begins not with welfare employment policy, but with new initiatives for child support. The Job Opportunities and Basic Skills Training program (JOBS) mandated by Title III of the act extends the equivalent of the WIN demonstration nationwide, but with many features missing from the programs we have studied. For example, JOBS requires participation of single parents with children no younger than three (one at state option); only five of the OBRA demonstrations had this feature, and of these none provide data specifically about the impact of the requirement on this group. JOBS provides funds for extensive education services; among the demonstration evaluations available to Congress in 1987, only Baltimore had this feature, and the consequences of the inclusion of education cannot be determined from the outcome. JOBS provides funds for extended Medicaid and day care for welfare recipients who become employed; none of the OBRA demonstrations that had been evaluated at the time the FSA was being formulated had this feature (although services of this type are provided in the ET-Choices and GAIN initiatives). As discussed earlier, the Family Support Act establishes a point-in-time participation target for JOBS eligibles; most of the OBRA demonstrations did not involve participation targets, and, as we have shown, for many it is difficult to assess just what participation indeed was. And finally, FSA mandates that states without AFDC-U adopt that program (with the option of limiting benefits to no more than six months during any calendar year) and, eventually, that all states require sixteen hours per week of CWEP participation for AFDC-U parents. While some experiments included AFDC-U families and required their participation in CWEP, almost no systematic study of this group is available in the OBRA evaluations.

Nonetheless, there exists a substantial consensus among persons active in welfare policy that the OBRA demonstrations, particularly MDRC evaluations of them, had a major effect on the course of the debate and, possibly, the success of the effort. In a review of MDRC's work that was commissioned by the Ford Foundation, Peter Szanton summarized a common theme in this way: "MDRC's work and welfare studies . . . were relevant, they were convincing, and they showed that something worked. They offered a solid basis for policy change and they helped to produce it" (Szanton, 1991, p. 600). In a similar vein, Ron Haskins, a minority staff member on the House Ways and Means Committee with responsibility for welfare-related legislation, writes: "Incorporating the best techniques of social science research, the MDRC studies showed that mothers could be helped into the labor force, that such results were impressive for precisely the group of mothers most likely to stay on welfare longest, and that such programs could be conducted on a cost-beneficial basis. As a result, the conclusions that states could conduct cost-beneficial job search and work experience programs did not give rise to the seemingly endless bickering among social scientists that accompanies most studies of welfare" (1991, p. 629). And finally, Dr. Erica Baum, Senator Daniel Patrick Moynihan's principal assistant for welfare policy at the time that the Family Security Act was prepared, writes that MDRC's work was a significant factor in the design and passage of the legislation: "MDRC's findings were unambiguous . . . [and] not subject to challenge on methodological grounds" (Baum, 1991, pp. 608-9).

As is surely evident from our discussions of the MDRC reports, we concur in much of the praise directed toward them. But as we have studied the outcomes, we have come to interpret the result somewhat at variance with the positions of Szanton, Haskins, and Baum. As we see it, the critical aspects of the research's impact on policy involve the image of process and what Baum has called the "sheer dumb luck" of timing.

Regarding process, as we discussed in our introduction, the foundation of much of the welfare reform effort of the 1970s had been the presumption that significant change required a "complete overhaul" of the welfare system. The alternative to radical change was incremental reform within the system, but in the words of the title of one paper on the subject, incremental reform was an idea "whose time had passed." A posture that called for radical reform had a number of corollaries. One was that there was not much to be learned about strategies for dealing with poverty from existing welfare institutions. A second was that since AFDC was a product of federal legislation, reform itself was a matter of national political initiative. The action was in Washington, and the focus of research in the various negative tax experiments was upon what were believed to be questions central to national income support policy. As Barbara Blum, the president of MDRC during the period in which the OBRA demonstrations were initiated, has pointed out, few state welfare administrators paid attention, or apparently were much aware of, the various negative tax experiments (Blum, 1989).

In contrast, the OBRA initiatives were specifically oriented toward state welfare operations. The theme repeatedly emphasized by MDRC was not that significant reform at the state level was infeasible--indeed, it was the very feasibility of such initiatives that MDRC wished to stress--but that if new initiatives were to have any impact, their effects would have to be evaluated in the style of the negative tax experiments. What was achieved was a sense both of feasibility of state-initiated policy modifications at existing welfare agencies and a sense that those changes which were feasible were worthy of laurels as reform. Moreover, the specifics were not as important as the concept of process: linking welfare receipt with obligation. States might construct their systems in a variety of ways, but for the most part such variation did not seem to violate common sense or compromise human dignity.

MDRC thus played both the role of messenger and medium, communicating the significance of appropriate research designs to the states and communicating findings from demonstrations that used appropriate designs to Congress.

Rationalization in the sense of integrating various programs pertinent to AFDC to better serve a common objective of increased recipient movement toward self-sufficiency had long been a theme of welfare reform efforts at the federal level. What MDRC seemed to be saying was that when given flexibility, states could achieve considerable rationalization of welfare systems by themselves. But it was not necessary to go as far as the Reagan administration had proposed early in the decade, that is, lumping virtually all poverty-related programs into a single block grant. Furthermore, given the platform of process, it was not necessary that all services be universal. Incremental reform, post-OBRA, meant enhancing the tool kit by providing states with a menu of services, some of which they could potentially adopt. And enhancing the welfare employment program tool kit is very much what the Family Support Act is about.

Timing was also essential to MDRC's impact. What Baum, Haskins, and Szanton all fail to note is that the perceived lack of ambiguity in MDRC's findings comes in significant part precisely because of timing. In 1987-88, virtually the only researchers familiar with MDRC's work were at MDRC. Results were communicated to Congress practically as they were developed. Had more time passed, it is likely that more criticism would have arisen, and, indeed, MDRC's findings would have been challenged on methodological grounds, just as we have done in this overview. More important, critics might have been able to focus better on the important distinctions between what was done in the OBRA demonstrations and what has been incorporated in JOBS. In 1987-88, however, fortuitous timing made MDRC results seem more compelling than they actually were. What Haskins fails to understand is that it requires time for social scientists to organize a good bicker!

Our Lessons

What judgments, then, are justified from the OBRA results? We feel a number of both negative and positive ones are appropriate.

The first is that, even with twenty-four observations, what went on in the OBRA demonstrations remains something of a "black box." Although in a number of instances we have what appear to be reliable estimates of the impacts of particular demonstrations, we are not very confident about what produced them, and we cannot predict with much certainty how the measured impacts would change if the demonstration were replicated in another location or if the demonstration program were expanded or made permanent in the same location. For example, moving from one location to another may change the level of participation, the intensity of treatment, the treatment mix, and the characteristics of the people receiving a particular treatment mix. This, in turn, may result in very different outcomes. Our suspicion that results from any specific OBRA demonstration may not be replicable is heightened by consideration of the special economic and political conditions that produced the demonstration data we now have.

Nevertheless, as stressed by participants in the FSA debates, it is reassuring that impacts from the demonstrations were usually in the desired direction--employment and earnings increased and (with somewhat less certainty) welfare dependency seemed to decline. However, these impacts were typically rather small in magnitude, although in interpreting their magnitudes, it is important to recognize that they were measured over both welfare recipients who actually received treatment under a program and welfare recipients who did not. It is not apparent from the comments by various participants that members of Congress or congressional staff ever grasped this distinction.

We are skeptical about attempts to glean information on subgroup effects from the OBRA research--for example, investigation of differential program effects on persons with long versus short welfare histories or persons with and without previous work histories. Even where differences in impacts do seem to exist across such groups, it is not clear whether they are attributable to different responses to the same treatment or to differences in the treatment actually received.

Findings from the benefit-cost analyses seem to be less reliable than those from the impact analyses. One reason for this is the inherent difficulty in measuring benefits and costs associated with complicated programs such as the ones tested. Some benefits and costs (e.g., the value of output produced in CWEP and net changes in the utility levels of program participants) are subject to potentially important biases. Moreover, benefit-cost analyses of some of the demonstration programs produced net gains for taxpayers and net losses for recipients, leaving us with what is basically a distributional quandary.

Truth can endure repetition, so we gladly repeat ourselves and echo what many others have said: policymakers owe a great debt to the Manpower Demonstration Research Corporation for imposing order on what almost certainly would have been chaotic in their (and the Ford Foundation's) absence. But a price has been paid. The MDRC approach to analysis is basically atheoretic, and the firm's approach to data has been proprietary. According to some observers, the monopoly established by MDRC's innovation served the public interest by clarifying issues and by focusing available information upon congressional concerns. But in the end, it is not clear that social science has been served. To our knowledge, not a single piece of research has been conducted by persons other than MDRC employees or consultants using primary data from the OBRA demonstrations collected by MDRC.

The problem arises, in part, from a factor that has contributed immensely to MDRC's success. The firm acts as an agent of the states with which it is working. This seems, among other things, to increase local willingness to assist and to motivate much greater interest in the evaluation findings than might otherwise be the case. But while it may be in California's interest to contract with MDRC for evaluation of GAIN, for most Californians, development of a public-use sample is not a major object of concern. Indeed, many contracts appear specifically to forbid the dissemination of data, and certainly funding is never provided for it. But we are not convinced that at least until recently MDRC sought funding for preparation of data for public use or that a major effort has been exerted with contractors to authorize such distribution. The challenge for national policy is to find ways to support firms like MDRC in their efforts (and to assure the motivation of their employees), while also assuring that as soon as possible the data upon which their results are based are made accessible and subject to the evaluation of the social science community at large. Of course, it is not MDRC's fault that the federal government did not choose to mount a systematic evaluation of the OBRA demonstrations, and the approach adopted by the firm may reflect the objectives of the foundations that have provided support. After all, none of the demonstrations we considered produced a public-use sample. But the government is involved in upcoming WEP evaluations, most notably of the JOBS program, and such matters must be considered. The evaluation plan for the JOBS program includes a public-use tape.

Future Research

We conclude this paper by providing a brief list of ways in which we believe future designs for evaluating employment-oriented programs for the welfare population could be improved.

(1) Site Randomization. It has been emphasized at several points in this paper that sites in the OBRA demonstrations were typically self-selected, thereby reducing the extent to which results from the demonstrations could be generalized. This problem might be ameliorated somewhat by random selection of sites in which to test various treatments. Although such an approach is probably infeasible at the state level, it would appear possible to randomly select welfare offices within a state as test sites.

(2) Analysis of individual treatment components. In all but a few of the OBRA demonstrations, those who were eligible were offered services in a package containing several individual

components. We believe that in future evaluations the interests of program operators as well as the larger policy community would be best served by emphasis on evaluation of specific components, perhaps in multisite demonstrations. Consequently, we strongly endorse evaluation designs in which program packages for which members of different treatment groups are eligible differ in terms of only a single program component.

(3) Evaluation of sanctioning. An important characteristics of the OBRA demonstrations is that most of the tested programs were mandatory. However, little was learned about the impact of sanctioning itself. We recommend that attention be given to design of alternative, more graduated, sanctioning procedures and that such graduations be tested in the context of a WEP operation. As long as the country is committed to obligation as a component of welfare policy, sanctions will be an instrument. Present procedures appear excessively cumbersome, and the sanctions available may be unnecessarily harsh.

(4) Modeling the selection process. We have emphasized throughout this paper that there seems to have been considerable variability in the treatment actually received among persons who were eligible to participate in any specific OBRA demonstration program, even when adjustment is made for the fact that some recipients stay eligible for only short periods of time. In future demonstrations, it would be very useful to investigate specifically how these treatment variations come about (for example, the role of the AFDC recipient versus the role of program administrators), and the association between such variation and client characteristics. Among other things, such study would allow detection of changes in service management procedures and, in consequence, treatments over time.

(5) Comparison of subgroup impacts. To allocate resources more efficiently, it is obviously important that we learn more about how treatment impacts vary among different subgroups. This, however, requires that members of different subgroups be treated as similarly as

feasible under the program being tested. We think it imperative that attention be devoted to the design of experiments, and the creation of incentives for governments to conduct experiments, in which discretion in service allocation is minimized and, in consequence, genuine service effects can be observed separately from the consequences of "creaming" or other administrative behaviors.

(6) Labor market intermediation. One of the most striking features of the OBRA demonstration results is the regular discovery of positive payoffs from what might at first blush seem relatively trivial interventions related to job search and employment preparation. To us, these research results confirm what traditional labor market economists have long argued: namely, that various features of labor markets pose significant barriers to low-productivity workers and that information both about jobs and about potential employee productivity is often hard to come by. More complex forms of assistance for workers moving into unsubsidized employment than those used in the OBRA demonstrations are available, including the efforts by some private firms to act as intermediaries between welfare offices and private employers. Investigation of the impacts of these innovations would seem justified, as would focusing attention on the consequences of increasing basic skills training.

(7) The research match. Part of MDRC's success is attributable to the financial support provided by the Ford Foundation. Currently, state funds spent on research on welfare innovations are typically matched with federal dollars on exactly the same basis as all other management expenditures. There seems to be no theoretical or practical justification for this. A more reasonable policy may be for the federal government to offer more generous funding for research on welfare program innovations in exchange for states meeting certain design and reporting requirements and directing evaluation efforts at important, unresolved policy issues--for instance, the consequences of individual WEP components for system outcomes. In its

sponsorship of MDRC's OBRA evaluations, the Ford Foundation created something of great value. But ultimately program evaluation is the business of government, and the problem of finding ways of getting this job done well by government deserves more attention.

Notes

[1]See Munnell (1986), especially the chapters by Burtless, Cain, and Hanushek.

[2]As Baum (1991) points out, this episode led many scholars to question the effectiveness of social science research in influencing policy. See Aaron (1978) and Glazer (1988).

[3]Although we focus on employment policy, keep in mind that an equally important theme of the Family Support Act is the enhancement of the obligation of absent parents to support their children. Here, also, a significant body of policy-related research and state innovation developed during the 1980s (Garfinkel and McLanahan, 1986). It would be interesting to compare the interactions among politics, research, and policy on the two fronts, but we do not attempt such a comparison here.

[4]Other situations conferring exemption include enrollment in school (for teenagers), illness, incapacity, lack of access to transportation or remote location, and responsibility for care of a sick person.

[5]The Family Support Act of 1988 required adoption of AFDC-U by all states; states without an AFDC-U program prior to 1988 were allowed to limit duration of receipt of welfare for families qualifying under this provision.

[6]See Plotnick and Lidman (1987) for a discussion of caseload forecasting.

[7]Blum (1989) discusses factors motivating state welfare administrators who agreed to evaluation of welfare employment programs by random assignment.

[8]Evaluation designs based on random assignment of sites (cf. the West Virginia AFDC-U experiment in Table 1) have the potential for detecting displacement effects. If displacement occurs, aggregate employment of regular public sector workers should grow more slowly in sites receiving the program treatment than in sites used as controls.

Bibliography

Demonstration Reports and Related Materials

Arizona

Sherwood, Kay E. 1984. Arizona: Preliminary Management Lessons From The WIN Demonstration Program. New York: Manpower Demonstration Research Corporation.

Arkansas

Friedlander, Daniel, Gregory Hoerz, Janet Quint, and James Riccio. 1985. Arkansas: Final Report On The WORK Program in Two Counties. New York: Manpower Demonstration Research Corporation.

Friedlander, Daniel and Barbara Goldman. 1988. Employment and Welfare Impacts of the Arkansas WORK Progam: A Three-Year Follow-Up Study in Two Counties. New York: Manpower Demonstration Research Corporation.

Quint, Janet with Barbara Goldman and Judith Gueron. 1984. Interim Findings From the Arkansas WIN Demonstration Program. New York: Manpower Demonstration Research Corporation.

California

San Diego I (Employment Preparation/ Experimental Work Experience Program)

Goldman, Barbara, Daniel Friedlander, and David Long. 1986. California: Final Report on the San Diego Job Search and Work Experience Demonstration. New York: Manpower Demonstration Research Corporation.

Long, David and Virginia Knox. 1985. "Documentation of the Data Sources and Analytical Methods Used in Benefit-Cost Analysis of the EPP/EWEP Program in San Diego." Unpublished technical paper, Manpower Demonstration Research Corporation.

Price, Marilyn, Daniel Friedlander, and David Long. 1985. Findings from the San Mateo County Employment Preparation Program New York: Manpower Demonstration Research Corporation.

Warren, Paul. 1985. An Analysis of Findings from the San Diego Job Search and Work Experience Demonstration Program. Sacramento, California: Office of the Legislative Analysis, California State Legislature.

San Diego II (The Saturation Work Initiative)

Hamilton, Gayle. 1988. Interim Report on the Saturation Work Initiative Model in San Diego. New York: Manpower Demonstration Research Corporation.

Hamilton, Gayle and Daniel Friedlander. 1989. Final Report on the Saturation Work Initiative Model in San Diego. New York: Manpower Demonstration Research Corporation.

Greater Avenues for Independence (GAIN)

Riccio, James, Barbara Goldman, Gayle Hamilton, Karin Martinson, and Alan Orenstein. 1989. The Greater Avenues for Independence (GAIN) Program: Early Implementation Experiences and Lessons. New York: Manpower Demonstration Research Corporation.

Wallace, John and David Long. 1987. GAIN: Planning and Early Implementation. New York: Manpower Demonstration Research Corporation.

Florida

Florida Department of Health and Rehabilitative Services. 1987. "Evaluation of the TRADE Welfare for Work Program." Report E-87-1. Florida Department of Health and Rehabilitative Services, Office of the Inspector General, Office of Evaluation and Management Review.

Illinois

Friedlander, Daniel, Stephen Freedman, Gayle Hamilton, and Janet Quint. 1987. Illinois: Final Report On Job Search and Work Experience in Cook County. New York: Manpower Demonstration Research Corporation.

Quint, Janet and Cynthia Guy, with Gregory Hoerz, Gayle Hamilton, Joseph Ball, Barbara Goldman, and Judith Gueron. 1986. Interim Findings from the Illinois WIN Demonstration Program in Cook County. New York: Manpower Demonstration Research Corporation.

Maine

Auspos, Patricia, George Cave, and David Long. 1988. Maine: Final Report on the Training Opportunities in the Private Sector Program. New York: Manpower Demonstration Research Corporation.

Maryland

Friedlander, Daniel, Gregory Hoerz, David Long, and Janet Quint. 1985. Maryland: Final Report on the Employment Initiatives Evaluation. New York: Manpower Demponstration Research Corporation.

Friedlander, Daniel. 1987. Maryland: Supplemental Report on the Baltimore Options Program. New York: Manpower Demonstration Research Corporation.

Maryland Department of Human Resources. 1987. "Final Evaluation of Maryland's Grant Diversion/OJT Demonstration Program." Baltimore, Maryland: Department of Human Resources, Office of Welfare Employment Policy.

Quint, Janet with Joseph Ball, Barbara Goldman, Judith Gueron, and Gayle Hamilton. 1984. Interim Findings from the Maryland Employment Initiatives Programs. New York: Manpower Demonstration Research Corporation.

Massachusetts

Behn, Robert D. 1989. The Management of ET Choices in Massachusetts. Durham, North Carolina: Duke University Institute of Policy Sciences and Public Affairs.

Garasky, Steven. 1990. "Analyzing the Effect of Massachusetts' ET Choices Program on the State's AFDC-Basic Caseload." Evaluation Review, 14(6): 701-710.

Massachusetts Department of Public Welfare, Office of Research, Planning, and Evaluation. 1986. An Evaluation of the Massachusetts Employment and Training Choices Program: Interim Findings on Participation and Outcomes, FY84-FY85. Commonwealth of Massachusetts.

Massachusetts Department of Public Welfare, Office of Research, Planning, and Evaluation. 1986. Follow-Up Survey of the First 25,000 ET Placements. Commonwealth of Massachusetts.

Nightingale, Demetra Smith, Douglas Wissoker, Lynn Burbridge, D. Lee Bawden, and Neal Jeffries. 1990. Evaluation of the Massachusetts Employment and Training (ET) Choices Program. Washington: The Urban Institute.

O'Neill, June. 1990. Work and Welfare in Massachusetts: An Evaluation of the ET Program. Boston: Pioneer Institute for Public Policy Research.

Minnesota

Minnesota Department of Human Services. 1987. The Community Work Experience Program in Minnesota. St. Paul, Minnesota: Minnesota Department of Jobs and Training.

New Jersey

Freedman, Stephen, Jan Bryant, and George Cave. 1988. New Jersey: Final Report on the Grant Diversion Project. New York: Manpower Demonstration Research Corporation.

North Carolina

North Carolina Department of Human Resources. 1985. "Final Assessment of the Community Work Experience Program Demonstration Project in North Carolina." North Carolina Department of Human Resources, Division of Social Services, Planning and Information Section.

Ohio

Potomac Institute for Economic Research. 1988. Impact of the Work Programs: A Long-Term Perspective. Washington: Potomac Institute for Economic Research.

Oregon

Oregon Department of Human Resources. 1988. "Section 1115 - Demonstration Project Grant No. 11-P-98080-10-05. Final Progress Report." Salem, Oregon: Oregon Department of Human Resources, Adult and Family Services Division.

Pennsylvania

Pennsylvania Department of Public Welfare. 1986. Evaluation of the Pennsylvania Community Work Experience Program. Harrisburg, Pennsylvania: Pennsylvania Department of Public Welfare.

South Carolina

Clarkson, M. H. Co., Inc. n.d. "Final Report on the Evaluation of the South Carolina Work Experience Program." Columbia, South Carolina: M. H. Clarkson Co., Inc.

Utah

Janzen, Frederick V. and Jeffrey Bartlome. 1986. "Emergency Welfare Work and Employment: An Independent Evaluation of Utah's Emergency Work Program." Final Report. Salt Lake City, Utah: The Social Research Institute Graduate School of Social Work, University of Utah.

Janzen, Frederick V., Jeffrey Bartlome, and Patrick Cunningham. 1987. "Emergency Welfare Work and Employment: An Independent Evaluation of Utah's Emergency Work Program." Final Report. Salt Lake City, Utah: The Social Research Institute Graduate School of Social Work, University of Utah.

Virginia

Price, Marilyn. 1985. Interim Findings from the Virginia Employment Services Program. New York: Manpower Demonstration Research Corportation.

Riccio, James, George Cave, Stephen Freedman, and Marilyn Price. 1986. Virginia: Final Report on the Virginia Employment Services Program. New York: Manpower Demonstration Research Corporation.

Friedlander, Daniel. 1988. An Analysis of Extended Follow-Up for the Virginia Employment Services Program. New York: Manpower Demonstration Research Corporation.

Washington

Fiedler, Fred. 1983. "Report: Intensive Applicant Employment Services Evaluation." Olympia, Washington: Washington State Department of Social & Health Services, Division of Administration & Personnel.

Nelson, Hal. 1984. "Evaluation of the Community Work Experience Program." Olympia, Washington: Program Research and Evaluation Section, Office of Research and Data Analysis, Division of Administration and Personnel, Department of Social and Health Services.

West Virginia

Ball, Joseph, with Gayle Hamilton, Gregory Hoerz, Barbara Goldman, and Judith Gueron. 1984. Interim Findings on the Community Work Experience Demonstrations [West Virginia]. New York: Manpower Demonstration Research Corporation.

Friedlander, Daniel, Marjorie Erickson, Gayle Hamilton, and Virginia Knox. 1986. West Virginia: Final Report on the Community Work Experience Demonstrations. New York: Manpower Demonstration Research Corporation.

Other OBRA-Related Evaluations

Cave, George. 1989. Subgroup Impacts of a Wage Subsidy Program: Randomized Trials of OJT for AFDC Recipients. Paper presented at the Annual Meeting, American Economic Association, Atlanta, Georgia, 30 December 1989.

Friedlander, Daniel. 1988. Subgroup Impacts and Performance Indicators for Selected Welfare Employment Programs. New York: Manpower Demonstration Research Corporation.

Friedlander, Daniel. 1989. Subgroup Effects of Large-Scale Welfare Employment Programs. Manuscript, October 29, 1989.

Gueron, Judith. 1986. Work Initiatives for Welfare Recipients: Lessons from a Multi-State Experiment. New York: Manpower Demonstration Research Corporation.

Gueron, Judith. 1987. Reforming Welfare with Work. New York: The Ford Foundation.

Gueron, Judith and Edward Pauly. 1991. From Welfare to Work. New York: Russell Sage Foundation.

Long, David and Virginia Knox. 1985. Documentation of the Data Sources and Analytical Methods Used in the Benefit-Cost Analysis of the EPP/EWEP Program in San Diego. New York: Manpower Demonstration Research Corporation.

Nightingale, Demetra Smith and Lynn C. Burbridge. 1987. The Status of State Work-Welfare Programs in 1986: Implications for Welfare Reform. Washington, D.C.: The Urban Institute.

Potomac Institute for Economic Research. 1988. "Impact of the Work Programs: A Long-Term Perspective." Washington, D.C.: Potomac Institute for Economic Research.

Schiller, Bradley E. and C. Nielson Brasher. n.d. "The Experience with Workfare: Lessons from the 1980s." Manuscript.

United States General Accounting Office. 1988. Work and Welfare: Analysis of AFDC Employment Programs in Four States. Washington, D.C.: United States General Accounting Office.

United States General Accounting Office. 1987. Work and Welfare: Current AFDC Work Programs and Implications for Federal Policy. Washington, D.C.: GAO.

General

Aaron, Henry J. 1978. Politics and the Professors: The Great Society in Perspective. Washington, D.C.: The Brookings Institution.

Barnow, Burt S. 1987. "The Impacts of CETA Programs on Earnings." Journal of Human Resources, 22(2): 157-93.

Baum, Erica. 1991. "When the Witch Doctors Agree: The Family Support Act and Social Science Research." Journal of Policy Analysis and Management, 10(4): 603-615.

Bassi, Lauri J. and Orley Ashenfelter. 1986. "The Effect of Direct Job Creation and Training Programs on Low-Skilled Workers." In Sheldon H. Danziger and Daniel H. Weinberg, editors, Fighting Poverty: What Works and What Doesn't. Cambridge, Massachusetts: Harvard University Press, 1986.

Blum, Barbara. 1989. "Bringing Administrators into the Process." Public Welfare, 48(4): 4-12.

Burghardt, John, and Anne R. Gordon. 1988. "The Minority Female Single Parent Demonstration: Description of the Local Context and Target Population." Report prepared for the Rockefeller Foundation. Princeton, N.J.: Mathematica Policy Research, Inc.

Burtless, Gary. 1989. "The Effect of Reform on Employment, Earnings, and Income." In Phoebe H. Cottingham and David T. Ellwood, editors, Policy for the 1990's. Cambridge, Massachusetts: Harvard University Press, 1989.

Burtless, Gary. 1990. "The Economist's Lament: Public Assistance in America." The Journal of Economic Perspectives, 4(1): 57-78.

Cave, George. 1987. "Sample Sizes for Social Experiments." Proceedings of the Survey Research Methods Section, American Statistical Association, 1987, 178-182.

Committee on Ways and Means, U.S. House of Representatives. 1989. Background Material on Programs within the Jurisdiction of the Committee on Ways and Means. Washington: U.S. Government Printing Office.

Doolittle, Frederick, Frank Levy, and Michael Wiseman. 1977. "The Mirage of Welfare Reform" (with Frederick Doolittle and Frank Levy). The Public Interest, 47: 62-87.

Garfinkel, Irwin and Sara S. McLanahan. 1986. Single Mothers and their Children. Washington, D.C.: The Urban Institute Press.

Glazer, Nathan. 1988. The Limits of Social Policy. Cambridge, Mass.: Harvard University Press.

Gramlich, Edward M. 1981. Benefit-Cost Analysis of Government Programs. Englewood Ciffs, N.J.: Prentice-Hall.

Greenberg, David H. and Marvin B. Mandell. 1991. "Research Utilization in Policymaking: A Tale of Two Series (of Social Experiments)." Journal of Policy Analysis and Management, 10(4): 633-656.

Greenberg, David H. and Michael Wiseman. 1992. "What Did the OBRA Demonstrations Do?" In Charles F. Manski and Irwin Garfinkel, editors, Evaluating Welfare and Training Programs. Cambridge, Mass.: Harvard University Press, 1992.

Gueron, Judith. Reforming Welfare with Work. 1987. Occasional Paper Number Two, Ford Foundation Project on Social Welfare and the American Future. New York: Ford Foundation.

Gueron, Judith. 1990. "Work and Welfare: Lessons on Employment Programs." The Journal of Economic Perspectives, 4(1).

Haskins, Ron. 1991. "Congress Writes a Law: Research and Welfare Reform." Journal of Policy Analysis and Management, 10(4): 616-632.

Heckman, James and Joseph Hotz. 1989. "Choosing among Alternative Non-Experimental Methods of Estimating the Impact of Social Programs: The Case of Manpower Training." Journal of the American Statistical Association, December.

Hershey, Alan. 1988. "The Minority Female Single Parent Demonstration: Process Analysis of Program Operations." Report prepared for the Rockefeller Foundation. Princeton, New Jersey: Mathematica Policy Research, Inc.

Ketron, Inc. 1980. The Long-Term Impact of WIN II: A Longitudinal Evaluation of the Employment Experiences of Participants in the Work Incentive Program. Wayne, Pennsylvania: Ketron.

Majone, Giandomenico. 1989. Evidence, Argument and Persuasion in the Policy Process. New Haven: Yale University Press.

Mathematica Policy Research, Inc. 1989. The Minority Female Single Parent Demonstration: Report on Short-Term Economic Impacts. Princeton, N.J.: Mathematica Policy Resarch.

Mishan, E. J. 1988. Cost-Benefit Analysis. 4th edition. London: Unwin Hyman.

Moffitt, Robert. 1991. "Program Evaluation with Nonexperimental Data." Evaluation Review, 15(3): 291-314.

Munnell, Alicia H., editor. 1986. Lessons from the Income Maintenance Experiments. Boston: Federal Reserve Bank of Boston.

Pearce, D. W. 1983. Cost-Benefit Analysis. 2nd edition. New York: St. Martin's Press.

Plotnick, Robert D. and Russell M. Lidman. 1987. "Forecasting Welfare Caseloads: A Tool to Improve Budgeting." Public Budgeting and Finance, Autumn.

Porter, Kathryn H. 1990. Making JOBS Work: What the Research Says about Effective Employment Programs for AFDC Recipients. Washington: Center on Budget and Policy Priorities.

Schram, Sanford and Michael Wiseman. 1988. "Should Families Be Protected from AFDC-UP?" Institute for Research on Poverty Discussion Paper no. 860-88, University of Wisconsin-Madison.

Smith, Ralph E. 1987. Work-Related Programs for Welfare Recipients. Washington, D.C.: Congressional Budget Office.

Szanton, Peter L. 1991. "'The Remarkable Quango': Knowledge, Politics and Welfare Reform." Journal of Policy Analysis and Management, 10(4): 590-602.

U. S. General Accounting Office. 1987. Job Training Partnership Act: Services and Outcomes for Participants with Differing Needs. Washington, D.C.: GAO.

U. S. General Accounting Office. 1987. Work and Welfare: Current AFDC Work Programs and Implications for Federal Policy. Washington, D.C.: U. S. General Accounting Office. Report GAO/HRD-87-34.

White, Halbert. 1980. "A Heteroscedastic-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity." Econometrica, 48(4): 817-838.

Wiseman, Michael. 1987. "How Workfare Really Works." The Public Interest, 89(Fall): 36-47.

Wiseman, Michael. 1976. "Public Employment as Fiscal Policy." Brookings Papers on Economic Activity, 1976:1, 67-114.

Wiseman, Michael. 1991a. "Research and Policy: A Symposium on the Family Support Act of 1988." Journal of Policy Analysis and Management, 10(4): 588-89 and 657-666.

Wiseman, Michael. 1991b. What Did the American Work-Welfare Demonstrations Do? Why Should Germans Care? Bremen, Germany: University of Bremen Centre for Social Policy Research. ZeS Arbeitspapier Nr. 9/91.