



Introduction

Burt S. Barnow
Johns Hopkins University

Thomas Kaplan
Institute for Research on Poverty
University of Wisconsin – Madison

Robert A. Moffitt
Johns Hopkins University

In August 1996, Congress and the President replaced the 60-year-old Aid to Families with Dependent Children (AFDC) program with a block grant, Temporary Assistance for Needy Families (TANF). The new block grant, enacted as part of the Personal Responsibility and Work Opportunity Reconciliation Act, allowed states to transform their public assistance programs in many ways and gave them great discretion within federal rules to do so. Some states, such as Kansas, responded guardedly to their new policy flexibility, making only the changes specifically mandated by TANF (Johnston and Lindaman, 1998). These changes were not trivial; the new law, for example, limited federal cash assistance to a total of five years for each adult and required at least 50 percent of single parents who received the assistance to work full time by 2002. Yet other states used their new policy flexibility to craft still more sweeping changes to AFDC. In many of the more aggressive states, requirements for school attendance, work, or an activity that leads to work, along with modified policies for

Evaluating Comprehensive State Welfare Reforms: The Wisconsin Works Program

adjusting family income according to family size, were bundled into packages designed to generate social messages that discourage dependency on public assistance, promote work, and influence family-formation decisions.

It has long been clear that Wisconsin would be one of the more ambitious states in welfare reform. An early draft of Governor Tommy Thompson's welfare initiative, Wisconsin Works (or W-2), was first circulated in the summer of 1995. The governor released his formal W-2 proposal in October of that year, nine months before the passage of the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA), which created the TANF program, and the Wisconsin legislature enacted W-2 into law in March 1996.

The W-2 legislation eliminated AFDC in Wisconsin and fashioned a replacement that was fundamentally and comprehensively different. Because the architects of W-2 believed that financial support should derive only from work, the new law requires W-2 participants to enter a work assignment immediately, with no grace period for the development and implementation of an employment plan. Because W-2 planners believed that clients should face the same constraints as the working poor, the new legislation includes immediate and sharp financial penalties for missed hours of work under W-2, makes families receiving W-2 child care assistance responsible for significant co-payments, and structures W-2 cash assistance so as not to vary by family size. Because program designers also held that W-2 participants should realize some benefits, as well as the constraints, available to the working poor not receiving public assistance, most W-2 participants under the new legislation receive all of the child support paid on behalf of their children (unlike the case with AFDC, in which all child support over \$50 per month was taken out of the AFDC grant). Moreover, families gain access to W-2 through an agency that provides general employment and labor exchange services, not through anything called, or designed to look like, a welfare office.

Although Wisconsin is not the only state engaged in such comprehensive reform, W-2 is certainly among the most ambitious of the current state initiatives. For several reasons described later in this introduction, and more fully in some of the chapters that follow, comprehensive reforms present special challenges to evaluators and may require new evaluation paradigms. Researchers at, and

Introduction

affiliated with, the Institute for Research on Poverty (IRP) at the University of Wisconsin–Madison, which had helped develop the current paradigms 25 years earlier through an evaluation of a negative income tax program in New Jersey, believed that the W-2 program offered a worthy test of their ability to meet some of the new challenges of evaluating comprehensive welfare reform. A working group of IRP researchers began holding regular meetings in the summer of 1995 in an effort to develop a plan for the evaluation of W-2. After several introductory sessions, a set of core issues was defined and each subsequent meeting concentrated on one of those issues. The discussions evolved into papers which were first presented to a national conference on the evaluation of comprehensive reforms held in November 1996.

This volume contains selected papers from that conference. All the papers have been revised with the aim of offering a broader audience a discussion of general issues in the evaluation of comprehensive state welfare reforms and an application of that discussion to the evaluation of welfare reform in one state. The volume illustrates the many challenges researchers will face in evaluating such programs and makes specific suggestions applicable to evaluations both in Wisconsin and elsewhere.

This introduction has three remaining sections. Because this volume is first and foremost about the methodology of evaluation, the next section offers an overview of the different types of program evaluations that can be conducted and of several of the key issues that arise in any evaluation. While these issues are familiar to evaluation experts, they may not be familiar to those outside the field and to government policy makers. The issues summarized in this section will also provide a context for the individual chapters that appear in the volume.

In the subsequent section, the details of the Wisconsin Works program are outlined. Because the chapters in the volume all apply their general principles to the W-2 program, it is useful to have an understanding of the program at the start.

In the final section, we summarize the contents and conclusions of each chapter in the volume.

Evaluating Comprehensive State Welfare Reforms: The Wisconsin Works Program

General Issues in the Impact Evaluation of Welfare Reform

The discussion in this section draws upon a long history of research on evaluation techniques. There are more detailed discussions of these issues in the many textbooks in the field (e.g., Rossi and Freeman, 1993). Our discussion pertains only to what is called “impact” evaluation, which is evaluation concerned with the effect of the program on the ultimate outcomes of interest, almost always outcomes at the individual or family level. (Chapter 2 contains a detailed discussion of outcome variables.) This type of evaluation is to be distinguished from “process” evaluation, which is concerned with assessing the effectiveness of the implementation of the program and in which, therefore, the outcomes of interest pertain to how well the program was put into place and its intentions realized in the field. Our discussion also explicitly addresses how each of the issues will have to be addressed in an evaluation of programs spawned by the PRWORA legislation, of which W-2 is one example.

Our overview addresses the following topics: (1) how to define the question of interest; (2) a classification of the different types of evaluation designs; and (3) data sources.

Defining the Question

In designing evaluations for social welfare programs, an evaluator must first determine the questions to be answered. Designs that are appropriate for answering some questions are often inappropriate for answering others. In addressing any particular set of hypotheses, there are five common questions to be considered:

- ❖ What is the intervention/activity of interest?
- ❖ What is the counterfactual to the intervention/activity of interest?
- ❖ What is the population of interest?
- ❖ What are the time frames of interest?
- ❖ What are the outcomes of interest?

For example, in many previous evaluations of welfare programs, the intervention studied has been a new job search, training,

Introduction

or employment intervention; the counterfactual has been an existing employment and training program or no program at all; the population of interest has consisted of volunteers or mandatory assignees to the "treatment" (that is, the new program or variation being tested) and has almost always consisted of individuals on the welfare rolls at the time of intervention (the general nonrecipient population has been excluded, by and large); the time frame of interest has typically run several years beyond the initiation of the reform; and the outcomes of interest have been the earnings, wage rates, and employment of the population as well as their receipt of welfare and amount of welfare payments. In a study of devolution resulting from PRWORA, many of these specifications are likely to change.

Activity/Intervention of Interest. Defining the activity or intervention of interest in the post-PRWORA environment raises many difficult issues. One is that the scope of the intervention must be much greater than anything tested in past evaluations. Devolution has already brought major changes to the AFDC program (including a new name), the food stamp and Medicaid programs, child care provision, and social services. Even if it were desirable, or possible, to focus solely on, say, the TANF program, the interventions already underway in most of the states are multifaceted and involve, most commonly, some combination of increased work requirements, heightened sanctions for failure to comply, time limits, financial work incentives, relaxation of AFDC-UP (Unemployed Parent) rules, family caps, and changes in asset requirements.

Disentangling the effects of interventions with multiple components creates what we term the "bundling" problem: The intervention is, in actuality, a "bundle" of different interventions, all of which are implemented more or less simultaneously. This will make it extremely difficult for an evaluation to estimate impacts of each of the individual interventions separately and to disentangle their separate effects. This problem can be overstated, however, because it is not clear that one should be interested in the impact of adding an individual component on top of the old program environment, or on top of the new environment after devolution. It is the argument of many advocates of devolution that the impact of the sum of the interventions is more than the sum of the impacts of each individually. To the extent this is correct, it is arguable that the first order of business is to estimate the impact of the bundle, and only secondarily to estimate the impact of the components; and it could

Evaluating Comprehensive State Welfare Reforms: The Wisconsin Works Program

also be argued that the impact of the components should be assessed against the new bundle rather than the old bundle.

Another problem is the infeasibility of conducting randomized experiments. In an experimental environment, it may be possible to test many of the major interventions with random assignment if it were desired to do so, although there are so many that the design matrix required would be very large. But if nonexperimental evaluations are the only options available, as we assume to be the case, then an evaluation must rely on “natural” variation — that is, on the choices that states and localities make. If this variation is very large relative to the number of states and localities, it may be difficult to infer impacts of individual components even if the impacts of bundled programs in a large number of states and localities are estimated.

The Counterfactual. The definition of a counterfactual is also problematic in the post-PRWORA environment because states and localities began implementing major changes in their programs before there was time to mount an evaluation. Consequently, it is too late, in a sense, to measure the “before” half of a before-and-after study. There are, however, two ameliorating factors in the problem of defining the “before.” First, to the extent that it is possible to gather information on past behavior from historical administrative data, it may be possible to establish a baseline prior to the current changes underway in the states. Second, it is quite likely that the implementation of devolution will take many years and, more specifically, that the changes in the program from 1999 to 2001, for example, may dwarf anything that has happened by the end of 1999. If so, there is still plenty of time to measure major impacts because relatively little has yet happened. This possibility also has implications for an evaluation design, for it implies that a good evaluation must maintain the flexibility to accommodate itself to an evolving policy environment.

Population of Interest. Most past welfare evaluations have had as their population of interest individuals on the welfare rolls or some subset of them (e.g., JOBS-eligibles). However, in addition to the fact that the impact of the new reforms is likely to be broader within the recipient population (i.e., to affect a greater proportion of recipients), it is likely to have major effects in the nonrecipient population as well through well-known “entry” effects. Most observers expect these entry effects to be negative, i.e., that there will be *deterrent* effects. This is often explicitly intended by the policy makers

Introduction

implementing the programs and is often regarded by those policy makers as a desirable outcome. The expansion of the population of interest calls for a much broader population base for the evaluation than has been the case previously. This population might be circumscribable to some extent, for example, by examining only the poor or near-poor population.^{1,2}

Finally, for some purposes it may be desirable to focus on particular subgroups of the population. For example, there is great interest in what effect time limits will have on AFDC recipients who exhaust their entitlement. It may be desirable to draw a special sample of near-exhaustees or to over sample such individuals.

Time Frames of Interest. There are several ways of considering the time frames of interest for the study. First, there is the issue of how long to track the sample or samples of participants. If time limits are included in the sites, it will be very desirable to track the research sample (or at least part of the sample) for at least one year beyond the point at which participants are terminated from the rolls. If five-year limits are imposed, this may be quite a long intervention. Rather different in spirit is the question of how long the time frame should be *prior* to the intervention examined. This is more a methodological question and we must postpone it to some extent to the discussion of alternative designs. One distinction worth making here, however, is that nonexperimental evaluations require, by and large, more historical data than do experimental evaluations because there is more need to control for the "histories" of the individuals and cities involved. In addition, as we noted previously, it may be desirable to construct sufficient historical data to get back to some defined "before" in any case.

Outcomes of Interest. At first blush, the outcomes of interest should be similar to the outcomes in previous studies: earnings, wage rates, employment, receipt of welfare, and amount of welfare received, for example. The broadening of the population we mentioned earlier, particularly to nonrecipients, requires that exit and entry rates be a part of the study of welfare receipt in this case. In addition, there may be interest in more aggregate outcomes such as caseloads and dollars spent on welfare.

The tradeoffs between different objectives may be more stark in the new environment than in the past, and it is worthwhile to be cognizant of that. By all appearances, the tradeoff between the goal

Evaluating Comprehensive State Welfare Reforms: The Wisconsin Works Program

of reducing the caseload, on the one hand, and increasing recipient well-being, on the other, is likely to shift more toward the former than has been the case in the past. Indeed, many policy makers are completely willing to accept reductions in recipient income and increases in poverty rates in exchange for caseload and cost reduction, particularly if it is achieved by altering the rules of the program in a way that is argued to be preferred by society. The notion of a time limit, for example, is partly based on the simple notion that only a certain amount of support should be given, regardless of the consequences.

Another implication of the current state-level interest in new “rules,” “norms,” and “expectations of behavior” is that implementation becomes more important, and is to some extent an outcome of interest in and of itself. If the major goal of current state-level policy makers is to require work while on the rolls and not to pay for more than five years of benefits or for additional children, and if these policies are aimed not at changing behavior but simply at enforcing what are asserted to be societal norms, then the goal of the program is merely to implement and enforce those rules, regardless of the consequences. That by itself may be difficult to do on a large scale, and it is not yet clear how successful states will be in that endeavor. But an evaluation design that ignores this point of view runs the risk of not answering the questions that some policy makers want answered.

A Classification Scheme for Nonexperimental Evaluations

Table 1 provides one type of classification scheme of the types of different evaluation designs, with a focus on the implicit counterfactual – that is, who is being compared to whom in order to obtain an estimate of the program effect. In nonexperimental studies, the estimation of a program, or “treatment,” effect is necessarily based on a comparison of different individuals or groups of individuals who face, or have been exposed to, different types of programs or program characteristics. We shall term these “quasi-experiments,” in line with traditional usage (Campbell and Stanley, 1966), in contrast to true controlled experiments. The table considers four different generic types of evaluations, and subcases within those: pure before-and-after designs, pure cross-section designs, designs which combine before-and-after with cross-sectional elements, and cohort designs. Each of the types of quasi-

Introduction

Table 1 Classification Scheme for Nonexperimental Evaluations		
<i>Generic Type of Study</i>	<i>Specific Type of Study</i>	<i>Description</i>
Pure Before-After	Individual Units	Units examined over time and outcomes measured; program has changed over time; attribute change in outcomes to change in program; can have multiple “before” and multiple “after” time periods
	Aggregates	Recipients or nonrecipients; in one area, most commonly; usually do not have a long time series; sometimes have subannual data and sometimes not Fixed geographic unit, usually a state; also called time series modeling; usually have relatively long time series and often have subannual data
Pure Cross-Section	Individuals Within Areas	Comparison of different units at a point in time (e.g., week, month, or year); program differs across units; attribute difference in outcomes across units to program differences
	Across Areas	Usually recipients only since recipient-nonrecipient comparisons usually not reliable; different individuals are treated differently; danger of selection bias Individuals or aggregates; danger of site effects
Cross-Section and Before-After	Individual Units Within Areas	Combination of two; have units that are treated differently; measure outcomes of all observations over time Different recipient or nonrecipients are treated differently; and treatment changes over time; permits “fixed effects” (or “differences in differences”) as well as “autoregressive” models that use lagged variables (“history”) to control for “heterogeneity”
	Individual Units Across Areas	Different recipients or nonrecipients in different areas with different programs, and program changes over time
	Aggregates Across Areas	Aggregate time series modeling but using multiple areas, usually states

**Evaluating Comprehensive State Welfare Reforms:
The Wisconsin Works Program**

Table 1
Classification Scheme for Nonexperimental Evaluations

<i>Generic Type of Study</i>	<i>Specific Type of Study</i>	<i>Description</i>
Cohort Design		Multiple birth or program entry cohorts who are each followed over time; program is changing over time; changes in cohort experiences are attributed to program change
	Individual Units Within Areas	Individual data on multiple cohorts within a single area
	Individual Units Across Areas	Multiple cohorts in multiple areas; can have cohort and area "fixed effects" by comparing cohort differences across areas
	Aggregates Across Areas	Possible if aggregate data can be disaggregated by age

Note: Each type can utilize administrative data, survey data, or both.

Introduction

experiments in Table 1 thus makes a different type of comparison. It would be a wonderful state of the world if they all generated the same impact estimates for the same type of program, but this is unlikely to be the case because the threats to the design are different in each case.

Pure before-and-after designs simply follow individuals or groups of individuals over a time period within which a program change has occurred. The change in their outcomes is attributed to the change in the program. The threats to this design are of two distinct types: aging (sometimes called maturation or life cycle) effects, and systematic external changes in the environment. Aging effects might be ignorable for short periods, but over longer periods the change in outcomes may be affected by natural life cycle patterns. In programs with short average duration, such as food stamps, aging effects can be important even for relatively short periods. Also important are changes in the local labor market, in the neighborhood environment, etc., which occur simultaneously with the program change and which therefore confound the measurement of its effects.

If aggregate data are used, either at a state level or national level, this approach is usually termed "time series modeling" or "caseload modeling." Aggregation has very little advantage per se, but aggregate data are often available for longer time periods and for more cross-section units (see next generic type) than are individual, micro data, which puts them at an advantage. Particularly in a before-and-after evaluation, where reliance on stability of the local economic environment is so important, a longer time series can be invaluable in separating the influence of general economic events from the program change in question. Aggregate data are also often significantly less expensive to obtain than micro data, and they can be used to detect entry effects.

These evaluation types can be further distinguished by whether administrative data, household survey data, or both are used. Aggregate data on caseloads or wages can be obtained from administrative data, and individual data may be gathered from welfare records or wage records, for example. Again, individual data are generally preferred, but they are also usually available for fewer time periods and fewer areas.

Evaluating Comprehensive State Welfare Reforms: The Wisconsin Works Program

The second type, pure cross-section quasi-experiments, are rare, even though they correspond most closely to controlled experiments. Comparing recipients to nonrecipients, different types of recipients, or different areas at a single point in time so obviously runs the risk of confounding program effects with other differences across individuals or areas that the method is almost never used. Instead, these types of comparisons are conducted when multiple periods of data are available, which falls under our third generic category of combined cross-section and before-and-after data.

This third generic category is, in fact, the most common type of nonexperimental evaluation and covers a large number of different subtypes. One classic method of evaluation is a comparison of participants to nonparticipants over time, using the history of their behavior to control for heterogeneity between the groups. However, while this method has a fairly long history in the evaluation of job training programs, it has almost never been used for the evaluation of welfare programs because welfare recipients and nonrecipients are generally thought to be sufficiently different to be noncomparable, even when observable histories are controlled for. Somewhat more common are comparisons among different types of recipients who are given different treatments (e.g., different employment and training programs), who are on waiting lists, or who are otherwise treated differently by the program. Fixed effects, autoregressive, matching, and other techniques are often used in this case to control for differences in histories between the groups compared.³ However, the main danger with these methods is that the differential treatment accorded to different groups of recipients or nonrecipients is endogenous (to use the econometric term); that is, that the different groups being compared differ along unobservable dimensions which make them noncomparable. With a few exceptions, therefore, although it is possible that credible and valid comparisons of this type might be found, most of the types possible run a serious risk of selection bias by assignment to category.

More credible are combination designs that compare recipients, nonrecipients, or both together, in different areas over time, thereby making use of the variation in program type between areas to measure program impacts. Aside from poor matching of initial conditions, the chief threat to this design is again uncontrolled differences in growth rates or other time-related changes in the outcome variables across areas, which is the counterpart to the

Introduction

unobserved site-effect problem in the pure cross-sectional comparison across areas. Comparison-site or matched-site designs are aimed at reducing this threat by choosing areas that are similar in a few measurable dimensions, but such designs are not successful if the areas differ in too many other ways. In addition, implementation of this method requires that the programs in different areas are capable of being compared along some common measure or scale, which may be difficult for current program changes, which are diverse and complex and very different across areas.

This method can be combined with use of historical data on individual recipients and non-recipients; there is no reason that one cannot, or should not, attempt to control for as much individual heterogeneity as possible even though cross-area comparisons of changes over time are the ultimate source of impact estimate. The comparison of different areas over time can also be achieved with aggregate data on caseloads, earnings, or other variables. Here again the chief advantage of such an approach lies in the greater number of time periods and areas available with aggregate data. However, as with the use of micro data, this method requires that programs in different areas and over time be ordered along some dimension or small number of dimensions. The development of a program typology would be one step in this direction.

The final generic method presented in Table 1 is a cohort design which measures program impact by comparing the experiences of different cohorts who face different programs because they come into contact with the system at different calendar times. A well-known example of this method is the evaluation of the 1981 Omnibus Budget Reconciliation Act (OBRA) legislation by the Research Triangle Institute (RTI, 1983). RTI compared the welfare exit and employment outcomes of a cohort of AFDC recipients prior to 1981 and a cohort after 1981; the latter experienced the OBRA legislation in full. The differences in outcomes between the two cohorts was attributed to OBRA.

This method can be extended in many ways. Multiple cohorts over time can be used for the comparison, which, if they are prior to the intervention, permits the incorporation of changes in the economic and social environment. Cohorts across areas can be compared, with the treatment measured as the across-area difference in cohort differences. Historical individual data on the individuals within each cohort can be collected and used as controls for

Evaluating Comprehensive State Welfare Reforms: The Wisconsin Works Program

heterogeneity. In addition, we should note once again that administrative data, household survey data, or some combination could be used for any of these designs.

The use of recipients or nonrecipients as separate groups in any of these quasi-experiments deserves special attention. While use of reciprocity as a defining characteristic is possible and desirable in many cases, it must be kept in mind that reciprocity itself is self-selected and may consequently pose a threat to designs which stratify on reciprocity. For example, comparisons of the outcomes of recipients in different areas or at different points in time within the same area, in order to draw inferences about the consequences of different programs or changes in programs over time, rely for their validity on the presumption that the population of recipients across areas or over time is the same in unobservable as well as observable dimensions. If this is not the case, differences in response may be the result of underlying differences in characteristics rather than of the differences in treatment. The same goes for comparisons of nonrecipients.

A reductionist position would be to only examine total populations of recipients and nonrecipients combined (or total populations of eligibles) across areas and over time. However, the gain in avoidance of selection bias has to be weighed against the loss of ability to pinpoint whether response is arising from recipients or nonrecipients, as well as having to rely on the alternative assumption that the two total populations are alike. This problem reflects a more general tension in nonexperimental analyses between a desire, on the one hand, to measure outcomes for narrowly defined subgroups of the population – not only for their intrinsic interest but also because it is easier to reduce the bundling problem, the narrower the subgroup – and the simultaneous desire to examine a large enough population to avoid separating the population into self-selected groups whose responses will differ because of unobservable differences in characteristics.

Data Sources and Issues

There are three general sources of data that can be used for the evaluation: aggregate administrative data, individual administrative data, and individual survey data. These forms of data are not simply substitutes, but can serve as complements as well – certain

Introduction

types of questions can only be answered with specific types of data. Thus, the question is not necessarily which type of data to collect for the evaluation, but which source or sources are needed to answer the questions of interest.

Aggregate Administrative Data. Aggregate administrative data consist of data collected by state and local governments for administrative purposes. Such data are generally available on a monthly basis and include total cases in the program, entries to the program, exits from the program (possibly by reason for exit), average benefit levels, and activity levels. They can be linked to demographic and economic data for the same area to estimate time series and determine the impact on caseload entries, exits, average number on the rolls, and benefit levels of program changes.

Aggregate data may be available only at the city or county level, but in some states it could be available at the office level. However, even if the caseload data are available below the city or county level, other determining variables that are needed for the analysis are likely to be unavailable below the local level. Thus, aggregate data are not likely to be helpful for any sub-city level analyses, for example.

Two advantages of aggregate data are that they are very inexpensive to obtain, and that they often go back for many years. In several states with which we are familiar, useful data go back to the early 1970s.⁴ These are both advantages in carrying out evaluations. The low cost implies that time series analyses can be used to supplement other, more expensive approaches. Aggregate data are often available for considerably longer periods of time than individual case record data are maintained, and the longer period means that evaluations can capture the effects of a wider range of economic conditions than if only a few years are available.

A problem that affects both aggregate and individual administrative data is that the analysis is limited by what has been collected. If, for example, one would like to include the number of single poor female heads of households as an explanatory variable in the analysis, that variable may not be available. Likewise, some outcomes of interest may not be available in aggregate administrative data.

Individual Administrative Data. This type of data includes information maintained at the individual level on program

Evaluating Comprehensive State Welfare Reforms: The Wisconsin Works Program

participation and individual characteristics. Examples include benefits received in programs such as AFDC, food stamps, unemployment insurance, and supplemental security income (SSI). Because participation in these programs is generally conditioned on income, we can also obtain information on earnings, wage rates, and hours worked.

The time period over which such data are available will vary from state to state and program to program. Some states, for example, routinely discard unemployment insurance wage records after three years. Confidentiality statutes may pose access problems in certain states and programs.

Although administrative data are less expensive to obtain than household survey data (see below), they are not cheap to obtain either. Many states' records, even when kept in machine-readable form, are poorly organized and require a significant amount of matching, merging, and file creation. Files are often "dirty" in the sense of containing many errors and discrepancies which the administrative agency has not spent the time to correct. In addition, if the data are not in machine-readable form but only in raw hard-copy form, there is an extra series of steps in preparing a format for the data and then entering it either by human coders or scanners. Errors can be introduced at this stage as well unless sufficient resources are allocated to keep them at a minimal level.

Leaving these practical difficulties aside, the use of administrative data has some limitations from a design point of view as well. First, by definition, a study using administrative data is limited to the variables available in the data – there is no flexibility to add or modify data elements. Administrative data tend to be rather weak on demographic information, for example. Second, administrative data can only track individuals while they are in the system. To know anything about earnings of families who leave the welfare rolls, for example, IRS records or unemployment insurance wage records must be obtained (although this is not uncommon in many evaluations). Third, these two problems also affect the ability to gather preprogram information on individuals in the evaluation, some of whom were off welfare.

Survey Data. Surveys of individuals of interest have several important advantages over the two other data sources, but they also have several important limitations. Perhaps the most important

Introduction

advantage of surveys is that data can be gathered on any topic amenable to survey questioning, and this broadens the types of information that can be obtained enormously. Besides the usual questions on outcomes and demographic variables, surveys can seek to cover topics not covered (or at least not covered in depth) in administrative data, such as motivation, mental health, intelligence, education, detailed work history, and so on. In addition, a survey can be designed to cover whatever population is of interest, although broad coverage may significantly increase the cost. Thus, a survey can generate information on a comparison group that would not be available from administrative data.

There are a number of important disadvantages of survey data as well, however. Perhaps the most significant is that they can be very expensive, particularly when the focus is a low-income population. Major costs may be incurred in developing a sampling frame, for example. In addition, screening costs can be extremely large when only a small proportion of the population is sampled in an area. In some instances these problems can be avoided, but they may reduce the validity of the survey.

Another major disadvantage of surveys is that they can only gather data prospectively. Because of recall problems, it is not possible to use a survey to cover previous time periods. Thus, if we wish to compare a program starting next month with the current program, we cannot rely on surveys to gather data on a previous cohort. This makes it difficult to address the problem of the "before" with survey data.

Conclusions

The most attractive designs are those that combine cross-section before-and-after variation and utilize cross-cohort variation. In any evaluation, it is important that the threats to the particular design chosen be examined and that the data collection plan take this into account. Supplementing individual administrative data with survey data, for example, can permit checks for self-selection into reciprocity; and supplementing an individual-level analysis with a time series analysis using aggregate administrative data can be used to check whether program effects are confounded with general trends. Although resource and time constraints obviously may limit

Evaluating Comprehensive State Welfare Reforms: The Wisconsin Works Program

such an ambitious data collection strategy, it should be regarded as a goal to be aimed for.

The Wisconsin Works (W-2) Program

Wisconsin's welfare reform plan, Wisconsin Works (W-2), is among the most aggressive of current efforts to "end welfare as we know it." The W-2 initiative, which eliminates AFDC and replaces it with cash assistance available only through work or participation in work-like activities, has several components. Participation requirements for low-income parents begin when the youngest child is 12 weeks old. The parents are assigned to a Financial and Employment Planner (FEP), who places them on one of four levels of a "self-sufficiency ladder" and helps and encourages them to move up the ladder to greater independence, as indicated on the grid in Table 2. The different levels have different grant amounts and also different time and participation requirements, as shown in the table. Small loans, which can be repaid in cash or community service, are available to help participants find and keep work. Two-parent families are eligible for all W-2 services if the families meet income and asset restrictions, although many services are restricted to the parent deemed most likely to be the primary earner. W-2 recipients in the two lower tiers technically receive a monthly grant, which drops by \$5.15 for each hour of failure to participate without good cause.

Unlike the AFDC program, the level of W-2 assistance does not depend on family size, but only on the case head's hours of participation and level on the W-2 self-sufficiency ladder. Also unlike AFDC, W-2 is not statutorily identified as an entitlement. Participation in the overall program is limited to five years, the maximum period for which the federal government will support most participants under TANF. Each level on the self-sufficiency ladder also has time limits, with extensions possible on a case-by-case basis. Moreover, W-2 eliminates the previous practice under which child support income beyond the first \$50 in a month goes to public agencies to reimburse welfare expenditures; most W-2 participants keep all child support paid on their behalf.

The W-2 program has a work-first emphasis and provides little support for formal education. The program does, however, offer child care and intends to offer health care assistance. The child care

Introduction

Level of W-2	Basic Income Package	Time Required of Recipients	Program Time Limits	Estimated Child Care Co-pays (\$/mo.)	
				Licensed Care	Certified Care
Unsubsidized Employment	Market wage + Food Stamps + EITC	40 hrs/wk standard	None	\$101-\$134	\$71-\$92
Trial Job (W-2 pays maximum of \$300/mo. to the employer)	At least minimum wage + Food Stamps + EITC	40 hrs/wk standard	Per job: 3 mo. with an option for one 3-mo. Extension; total 24 mo.	\$55	\$38
Community Service Job (CSJ)	\$673 per mo. + Food Stamps (no EITC)	30 hrs/wk standard; and up to 10 hrs/wk in education and training	Per job: 6 mo. with an option for one 3-mo. extension; total: 24 mo.	\$38	\$25
W-2 Transition	\$628 per mo. + Food Stamps (no EITC)	28 hrs/wk work activities standard; and up to 12 hrs/wk in education and training	24- mo. limit, but extensions permitted on a case-by-case basis	\$38	\$25

Sources: K. F. Folk, "Welfare Reform under Construction: Wisconsin Works (W-2)," *Focus* 18, no. 1 (special issue 1996): 55-57, and presentation materials created by the Wisconsin Department of Workforce Development.

Note: Estimated child care co-payments are for a three-person family with two children in care and receiving no child support payments. To estimate child care co-payments, the Trial Jobs position is assumed to pay minimum wage (\$5.15 per hour, or \$858 per month), and the pay for unsubsidized employment is assumed to range from \$6-\$7 per hour, or \$1,000-\$1,170 per month.

Evaluating Comprehensive State Welfare Reforms: The Wisconsin Works Program

benefit requires a participant co-payment, the level of which is based on family income and on the number of children in care. The co-payments are structured so as not to exceed 16 percent of family income and to be 30 percent lower for child care that has received county-level "certification" than for child care fully licensed by the state.

Wisconsin has also requested approval of a federal waiver to expand the current Medicaid program in a way that would serve most W-2 participants. Under the waiver request, submitted formally in January 1998 and approved to begin on July 1, 1999, the state would offer a new form of insurance, called BadgerCare, which would provide the state's current Medicaid benefits to a broader population, including all W-2 participants who do not have employer-paid insurance. Eligibility for the program extends to families with incomes up to 200 percent of the federal poverty line; those with incomes above 150 percent of the poverty line would pay a monthly premium of 3 percent to 3.5 percent of family income.

In general, the W-2 program reflects efforts by program planners to adhere to six core principles:

1. Parents without a disability should work, and should obtain no entitlement to cash assistance in the absence of work.
2. Expectations for success in the labor market should be high; custodial parents will live up or down to the expectations imposed upon them.
3. All cash benefits should be time-limited.
4. Government programs should provide child care and health care assistance to the working poor, defined in W-2 as families with incomes up to 200 percent of the federal poverty line, not just to public assistance recipients. (W-2 grants and employment assistance are available only to families with incomes below 115 percent of the poverty line.)
5. Those who receive grants and other benefits should face the conditions that affect the working poor: grant recipients should have to work; their first grant check should come only after a period of work; workers who receive public child care assistance and health care insurance should have to pay part of the

Introduction

cost of their benefits; and program participants should receive all child support paid on behalf of their resident children.

6. Competition to meet selected outcome criteria, not government monopolies, should determine who administers public assistance. Key administrative choices ought not rest upon traditional relations between the state and its counties or traditional conceptions that only public employees should have access to sensitive information and control program benefits.

Although comparative judgments concerning TANF programs nationally must be imperfect, owing to the difficulty of understanding all of them in sufficient detail, at least three key features of W-2 are unusual and perhaps unique among state TANF programs. First, except for W-2 participants in the lowest tier of the program, the only income available is through work, and financial penalties for failure to work start immediately. Second, W-2 contains less social contract language in which the state and the public assistance recipient agree on reciprocal obligations — the one side to make opportunities available and the other to pursue those opportunities — than is the case in many states. The difference between Wisconsin and some other states in this regard is subtle but meaningful. W-2 certainly provides help to program participants, especially with child care and health care. Moreover, the administrative rules for W-2 require local agencies “in consultation with the W-2 participant [to] develop a written employability plan for a W-2 participant which includes the participant’s W-2 employment position placement, required activities . . . and an unsubsidized employment goal.” The state’s W-2 policy document also allows FEPs to excuse participants from work requirements if child care is unavailable. But unlike many states, the FEP has complete discretion to make this determination of child care unavailability and complete authority to determine whether a participant needs a subsidized job. An applicant can appeal a decision of the FEP to the W-2 agency, which must rule on the FEP’s decision within 45 days. Even if the appeal is successful, however, the applicant receives no back payment for the appeal period (Wisconsin Department of Workforce Development, 1997, p. iv-18). Although some FEPs may alter these emphases in their daily practice, the primary focus of W-2 is on the participant’s obligations to follow the employability plan or, if considered ready for an unsubsidized job, to secure one. The emphasis is not on the responsibilities of the state or the W-2 agency to find them a job or train them for emerging opportunities.

Evaluating Comprehensive State Welfare Reforms: The Wisconsin Works Program

The third atypical feature of W-2 is its treatment of child support. With the exception of 4,000 families included in a control group for evaluation purposes, all other W-2 families will be able to keep all child support paid on their behalf. This marks a change from the policy that existed under AFDC, when a family could keep only \$50 per month of child support paid on the family's behalf. Any additional child support reimbursed the state and federal governments for their AFDC expenditures. Wisconsin is the only state that has chosen to pass through all child support to the resident parent. In fact, 30 states plus the District of Columbia have used their new flexibility under TANF to move in the opposite direction, keeping for the government all child support paid on behalf of a TANF family (U.S. Department of Health and Human Services, 1997).

The Chapters in This Volume

The chapters in this volume address both general evaluation issues and specific applications to the W-2 program. The initial chapters by Corbett, Cancian and Wolfe, Cain, and Haveman deal with fairly general issues, and the chapters that follow address W-2 more specifically.

In his beginning chapter, Thomas Corbett provides an overview of the evaluation problem. Corbett stresses that the evaluation problem in the current era of welfare reform will challenge traditional evaluation methods. He stresses that conventional impact analyses will be difficult to conduct because experimental methodologies will be impossible and nonexperimental methods will be, as always, subject to multiple interpretations. Moreover, he points out that process and implementation will be more important to welfare reform evaluation than they have been in the past. The reduced role of the federal government in evaluation is another significant feature of this round of welfare reform, according to Corbett. This implies that the varieties of evaluation strategies across states will be much greater than in the past and there is some danger of inadequate evaluation in some states. Finally, Corbett notes that the emphasis of the new round of welfare reforms on changing individual behavior provides another new, major challenge.

Corbett provides a useful history of the way in which the waiver process developed in the 1980s and the role of evaluation in

Introduction

that process. He then goes into great detail on the types of attributes that new welfare reforms are taking, and how they differ markedly from those in past reform efforts. Multiple characteristics of the welfare system are changed at the same time, and complexity will increase as different participants are treated differently. These developments make evaluation difficult, for they leave unclear the most desirable counterfactual, the appropriate unit of analysis, the target group of interest, the role of implementation analysis, and whether it is desirable or undesirable to capture local diversity and discretion in the evaluation. Corbett concludes by recommending the development of a common set of standards for evaluation, that expert review panels be formed, and that a strategy for diffusing information be developed.

In their paper, Cancian and Wolfe discuss the variety of different outcomes that might be considered in a welfare program evaluation and what constituencies exist for different types of outcomes, and they outline a number of trade-offs that arise in making these choices. Constituencies who have an interest in welfare reform outcomes include the affected families themselves, citizens, government officeholders, and program administrators, but also private charities, employers, schools and teachers, and members of the medical community. Because different constituencies have different interests, Cancian and Wolfe show that several trade-offs will arise. One trade-off is between studying a fairly narrow population that is served by the program vs. a broader population that is potentially affected. Many of the issues here are data-oriented because broad samples come from different data sets than those with necessary numbers of program participants. A second trade-off is between individual vs. community-based analysis. One example is that a new TANF program may have community-wide impacts that extend beyond the particular individual recipients affected. A third trade-off is between evaluation and monitoring, for evaluation is a causal inquiry which requires a well-defined counterfactual, whereas monitoring requires only collecting data on individuals over time. A fourth trade-off is that between a short and long time frame. Some income and labor market effects may occur quickly, but some effects (e.g., those on children) may occur more slowly.

Cancian and Wolfe apply these trade-offs to a consideration of four specific outcomes: work requirements, child care, child health, and family formation. For each they delineate a specific set of outcomes to be examined and whether each applies to a narrow or

Evaluating Comprehensive State Welfare Reforms: The Wisconsin Works Program

broad population. They then discuss the tradeoffs between individual and community-level study, between evaluation and monitoring, and between short and long time frames for each. Cancian and Wolfe conclude that priorities will have to be established given limitations in resources, and they offer the view that priority should be given to outcomes that can be simply monitored instead of evaluated, to outcomes that can be measured with administrative data, and to broad population outcomes that can be examined with existing national-level data sets.

Turning to the issue of evaluation methodology, Cain considers the usefulness of controlled experimentation in the new era of welfare reform. He begins by considering a traditional experiment with randomized experimentals and controls which tests the effect of a specific training program on welfare recipients. Two shortcomings of such experiments are that they would not detect entry effects and that they would not detect market-wide effects. Despite these drawbacks, such an experiment would be useful if the programs, populations, and outcomes are narrow and well-defined. However, the new welfare reforms are different and pose two major problems for controlled experiments, according to Cain. One is that the appropriate counterfactual of interest to policy makers is unclear, for the old welfare system is clearly not the alternative of interest, but it is not clear what other bundle of policies is. Another is that the effects of the new welfare are likely to be so broad in terms of markets, program services, and culture that any small-scale experiment will miss important outcomes and will yield inaccurate answers because the controls are likely to be affected. Cain points out other important issues in experiments that would create problems, including ethical issues, a sufficiently long time frame, the problem of continually evolving programmatic environment, and the lack of a clear-cut "before." Cain concludes that controlled experiments are not a practical method for evaluating the new welfare reforms.

Cain proposes instead that evaluators consider gathering longitudinal data across multiple states with different programs, while acknowledging the potential problems with cross-state comparisons. He proposes specific ways to control for confounding cross-state differences and to utilize the variety of programs implemented in different states. Cain ends up recommending evaluation within the traditional framework of a theoretically based nonexperimental study.

Introduction

Robert Haveman investigates alternative designs for evaluating national welfare reforms as well. Haveman begins by noting the characteristics of the new welfare reforms that will create difficulties for national program evaluation — the fact that every state can design its own program, that state-level conditions will be changing at the same time as programs are implemented, that the speed of implementation will vary across states, and that the scale of the welfare within states is very large. Haveman then lays out the features of an “ideal” evaluation and develops basic principles for securing reliable evaluations. The latter include the need to specify precisely the counterfactual to the policy and that the outcomes under both policy and counterfactual be measurable. Haveman then uses these features and principles to consider experimental designs, comparison site designs, and pre-post designs. He concludes that experimental designs have a fatal flaw when applied to policies that change the culture in a state, for controls will be affected by that change as well. He also concludes that comparison site designs will run the risk of comparing noncomparable states, and that because all states have a “current policy” in effect, none provides an appropriate counterfactual. The pre-post design, while having the problem of separating program effects from other “state-of-the-world” changes, at least clearly identifies the with- and without-program groups and clearly identifies different policy regimes, or states of the world. Haveman then specifically considers the Wisconsin Works program and concludes that the pre-post is most feasible of the evaluation options, in part because the limitations of the experimental and comparison site designs are even greater. He ends his chapter by proposing a specific data collection strategy for a pre-post evaluation of the W-2 program.

The remaining chapters in the volume turn specifically to the Wisconsin Works program. The first chapter in this series, by Kaplan and Meyer, proposes in great detail a specific evaluation plan for the W-2 program. Kaplan and Meyer propose an evaluation built around six central impact domains — income, dependency, child care, child welfare, health status, and living arrangements and family structure — and propose a basic evaluation that focuses on individuals rather than other groups. Like Haveman, they propose a pre-post design and they also propose to rely primarily on administrative data for cost reasons, including tax, unemployment insurance earnings, W-2, and food stamp records. Then, for each of the six domains, they consider: (1) the likely effects of W-2 on those outcomes, (2) what data, comparison group, and analysis approach

Evaluating Comprehensive State Welfare Reforms: The Wisconsin Works Program

should be used, and (3) what issues and limitations would arise. For example, for the study of the impact of W-2 on income, they propose a variety of comparisons of W-2 and AFDC recipients before and after the implementation of W-2, both new entrants and current participants, and to compare the incomes of these groups for several years (up to seven) after the date they are observed on welfare. Thus, for example, they would compare incomes of 1988 AFDC recipients from 1989 to 1995 to the incomes of 1998 W-2 recipients from 1995 to 2005. To examine entry effects, they propose to draw two similar samples of low-income families based on Wisconsin tax records and to follow them in the same manner as the recipient samples. They then consider specific data issues such as the definition of income, definition of W-2 recipient, and what variables to hold constant in the analysis. They then conduct a similarly detailed treatment of the other five domains of interest. Kaplan and Meyer also consider several secondary outcomes, such as homelessness, residential mobility, health insurance coverage, and effects on children with disabilities.

Four papers on process analysis and data issues follow. The first, by Corbett and Boehnen, focuses on process analysis. Corbett and Boehnen distinguish process analysis from other types of evaluation and then consider the components of a process analysis, including a description of the implementation of the program, an analysis of the discrepancy between intended and actual sequences of activities, a participation analysis, as well as analyses of dosage or intensity, continuity of program engagement, program coherence, cross-site comparisons, and overall quality. They then outline the key institutional actors in the W-2 program at a very detailed level, down to the receptionist who performs gatekeeping functions, and the key procedural steps in the program, including signaling, gatekeeping, triage/plan development, participation/monitoring, review/adjustment, and exit/follow-up. Corbett and Boehnen also show the client and participant service flow in W-2. They then turn to local variation in implementation in Wisconsin and to the key problem of measurement of participant experiences and participant flow, and discuss a variety of data sources that might be used for such measurement.

In the next chapter, Holden and Reynolds also discuss process analysis. They list three reasons that a process analysis is an essential supplement to an "outcomes" evaluation: to validate that program services were in fact delivered in the way assumed in the

Introduction

outcome analysis, to determine why a program did or did not work according to the outcome analysis, and to promote replication of the program in the future and therefore the utilization of the evaluation results for future programs. Holden and Reynolds also list key specific questions that should be addressed in a process analysis, including whether administrative services were in fact in place, whether services were delivered to the appropriate target population, whether there was variation in program services and administration across sites, whether the delivery of program services changed over time, and whether program services can be validly linked to program outcomes. The authors then turn to a discussion of the W-2 program and stress the complexity of the program services and that service delivery is very likely to vary substantially across areas within the state. Holden and Reynolds list the processes that will be important to the outcomes studied by other papers in this volume and then discuss the data necessary for assessing process. They conclude that the data gathering for a process analysis of W-2 will be the most challenging part of the task.

The chapter by David deals with the issues in measuring income for the W-2 population and for a non-W-2 population with administrative records. David proposes to match the W-2 administrative data base with Wisconsin tax record administrative data for a basic data set. He then assesses the utility of these sources of cross-section and panel data on income and other characteristics and shows how a matched beneficiary data-tax return panel could be constructed from Wisconsin records.

The last five chapters in the volume consider different specific outcome domains in a W-2 evaluation. The first, by Sandefur and Martin, discusses alternatives for collecting and analyzing data on family structure, maternal health, and child health in the context of evaluating W-2. The authors first develop hypotheses regarding how W-2 should affect the outcomes of interest. Based on the changes from AFDC to W-2, they hypothesize that W-2 should lead to reduced out-of-wedlock childbearing; they conclude that the impact of W-2 on maternal and child health is ambiguous in direction. Much of the data that would be required for an evaluation on the topics discussed by Sandefur and Martin can be obtained from administrative data sources, but they conclude that several administrative data sources would have to be merged, that some additional data would have to be collected through surveys, and that the

Evaluating Comprehensive State Welfare Reforms: The Wisconsin Works Program

National Longitudinal Survey of Youth (NLSY) includes many of the questions that would be needed for the evaluation.

Karen Folk and Marianne Bloch discuss how the child care services under W-2 can be evaluated. Folk and Bloch begin by noting that child care has been shown to be the greatest barrier to the employment of mothers of young children. They go on to note that an unusual feature of regulated child care in Wisconsin is the presence of a three-tier system of child care — licensed, certified, and provisionally certified providers. Folk and Bloch predict several effects of W-2 child care provisions. First, the strict work requirements of W-2 should lead to a large increase in the demand for child care that exceeds supply, especially in the short run. Second, the W-2 co-payment schedule may lower net income for some families relative to pre-W-2 levels. Third, the multi-tiered system can lead to changes in the quality of child care provided, but the direction will vary among families. The authors discuss the uses of various forms of administrative data, but they suggest that longitudinal data starting prior to implementation of W-2 will be required to fully answer the questions of interest. The more qualified licensed care facilities cost W-2 families more than certified facilities, so the W-2 recipients have an incentive to use the lower-quality care.

Daniel Meyer, Maria Cancian, and Emma Caspar describe the evaluation of child support under W-2 that they are undertaking. Wisconsin is evaluating the new policy under an experimental design in which the control group receives a smaller pass-through of child support. Because this component of the W-2 program is being evaluated with an experimental design, some of the problems encountered in evaluating the other components of W-2 are not present in the child support evaluation. They note, however, that some of the outcomes may take a long time to take effect. The authors note that outcomes of interest include formal child support, informal child support, paternity establishment, and earnings of both the custodial and noncustodial parents.

George Jesien, Caroline Hoffman, and Thomas Kaplan analyze the potential effects of W-2 on children under age three with developmental disabilities. Jesien and his colleagues note that having a child with developmental disabilities places additional demands on families' financial, time, and emotional resources. Of the estimated 7,000 families in the state with a child under three with a diagnosed disability, 2,250 will be required to meet the work requirements of

Introduction

W-2. The authors suggest outcomes of interest include the ability of the family to remain together, the ability of parents to participate in the services and programs for their disabled children, and the health and well-being of the disabled children subject to W-2. The availability of good records on disabled children prior to the implementation of W-2 might permit a pre-post approach to an evaluation, but the authors state that such an evaluation would be complicated by other changes that occurred about the same time that W-2 was implemented, such as tightening of eligibility for supplemental security income (SSI) because of the Supreme Court's Zebly decision. The authors suggest three approaches to the evaluation: case studies on a number of families with eligible children, a panel study to trace developments of selected families with children with disabilities, and a review of administrative records.

The chapter by Mark Courtney discusses how to evaluate the child welfare services component of W-2. Courtney notes that W-2 is likely to have mixed financial effects on families needing foster care and related services. Some families will benefit from the greater earnings induced by the program, but others may lose income if they cannot or choose not to meet the work requirements of the program. Courtney raises some of the difficult issues that must be faced in evaluating the welfare services component of W-2. He notes that use of out-of-home care is one outcome of interest that should be easy to measure, but maltreatment of children, another outcome of interest, will be much harder to capture. Courtney then weighs the strengths and weaknesses of alternative evaluation strategies. Administrative data often have many advantages, but in this instance pre-W-2 administrative data on foster care entry and child maltreatment are not available. Given the data limitations, Courtney recommends that a panel study of W-2 children be initiated, although given the rarity of some of the outcomes of interest, the sample would have to be quite large to provide meaningful results. Finally, Courtney suggests that experimentation at the county level be considered.

References

Campbell, D. T. and J. C. Stanley. *Experimental and Quasi-Experimental Designs for Research*. Skokie, IL: Rand-McNally, 1966.

**Evaluating Comprehensive State Welfare Reforms:
The Wisconsin Works Program**

Johnston, J. M. and K. Lindaman. "Implementing Welfare Reform in Kansas: Politics and Administration." Paper presented at the Annual Conference of the Midwest Political Science Association, April 23, 1998.

Research Triangle Institute (RTI). *Final Report: Evaluation of the 1981 Amendments*. Research Triangle Park, NC: Research Triangle Institute, 1983.

Rossi, P. and H. Freeman. *Evaluation: A Systematic Approach*. Newbury Park, CA: Sage, 1993.

U.S. Department of Health and Human Services, Office of Child Support Enforcement. *Child Support Report*. Washington, DC: Department of Health and Human Services, December 1997.

Wisconsin Department of Workforce Development. *Wisconsin Works Policy*. Madison, WI: Department of Workforce Development, 1997.

Introduction

Endnotes

- 1 To simplify the discussion, we ignore the AFDC-UP program here.
- 2 For some approaches, such as caseload modeling, it is not necessary to establish a population base for drawing the population of interest.
- 3 The method of “selection bias modeling,” associated with a certain tradition in econometrics, is sometimes associated with these methods as well. However, it is in fact not a separate method from any of those presented in the table, each of which could be formulated as a “selection bias model.”
- 4 The period that can be analyzed depends on more than just the availability of data. For example, New Jersey officials warned us that data prior to 1978 were of uncertain quality. Also, if the program structure changes in major ways, it may not be advisable to assume that the same model applies before and after the change. For example, the changes instituted by OBRA in 1981 may make it inadvisable to use pre-OBRA data in some states.

