

2

Research on PRWORA: What Can Be Learned from Four Large-Scale Projects Currently Underway

*Peter H. Rossi**

This paper assesses four major research projects currently underway, each designed to gather information on what is happening to low-income families subsequent to the welfare reforms instigated by the passage of the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA). Each of the projects is planned to describe some aspects of the welfare system in some places in the United States and the conditions of low-income households before and, in some cases, after the provisions of PRWORA have gone into effect. Although data collection has begun and the broad outlines of the research designs have been laid out, all the projects are still in progress. Changes in plans are likely to be made, especially for data collection efforts, which have not yet begun. Some data have been collected in each project, but the data collection will not be completed for several years. And, of course, except for the waiver experiments, none of the data collected has been analyzed in detail, although some findings have been released.

This review is based on only a few published documents: Most of them are unpublished memoranda, proposals, drafts of papers, and questionnaires intended to be used in the field. *Accordingly, this review must be regarded as a description of work still in progress.* Of course, a final report reviewing findings can only be attempted some years hence, when data collection will be complete and analyses will have been written.

The five research organizations involved—the Bureau of the Census, Manpower Demonstration Research Corporation, the Urban Institute, Mathematica Policy Research, and Abt Associates—generously made available the materials on which this review is based. I am especially grateful to Daniel Weinberg and Michael McMahon (Bureau of the Census), Gordon Berlin and Charles Michalopoulos (Manpower Demonstration Research Corporation), Fritz Scheuren,

*Peter H. Rossi is S. A. Rice Professor Emeritus at the University of Massachusetts (Amherst).

2: Research on PRWORA: What Can Be Learned

Genevieve Kenney and Anna Kondratas (Urban Institute) and Howard Rolston (Department of Health and Human Services), all of whom sent me materials and patiently answered my questions. I circulated earlier versions of this paper to them, and their comments helped correct errors in my descriptions of the projects. I did not always agree with some of their judgments, but I did take them into account in this revision.

EXECUTIVE SUMMARY

America's public welfare system "as we knew it" changed dramatically with the passage of the Personal Responsibility and Work Opportunity Reconciliation Act of 1996. The major change was that Aid to Families with Dependent Children (AFDC) was abolished and replaced by Temporary Assistance for Needy Families (TANF), which mandated a five-year time limit on federally funded assistance, new work requirements, and a cap on federal funds based on historical AFDC funding. Otherwise, it is a program to be designed in detail and run by the states, emphasizing employment, marriage, and reduced out-of-wedlock childbearing.

Although few tears were shed over the death of AFDC, both advocates and opponents of PRWORA became concerned about how the new law, especially the TANF provisions, would affect low-income families. PRWORA advocates worried whether TANF would reduce welfare dependency and move families off the welfare rolls as intended. Opponents worried that low-income families would be plunged into abject poverty and that the children of the poor would suffer greatly. Fueled by these concerns, several major research projects were started and ongoing projects modified in order to provide empirical data on the changes wrought by TANF and other PRWORA-mandated changes and to estimate the new program's impact.

This paper assesses the prospects of each of four major research projects to provide empirically based findings concerning how well low-income families fared before and after PRWORA, how those prospects varied from state to state, and whether those trends can be credibly attributed to PRWORA.

The four research programs are as follows:

- **The Survey of Program Dynamics.** The Bureau of the Census is extending the Survey of Income and Program Participation (SIPP) to create the Survey of Program Dynamics (SPD), a longitudinal survey consisting of repeated interviews of a national sample of households from several years before PRWORA to several years after enactment. SPD oversamples low-income households.
- **Assessing the New Federalism.** The Urban Institute's extensive study, *Assessing the New Federalism*, is examining the devolution shifting many welfare programs from the federal to the state level. Administrative studies will describe the changes in health and welfare systems in each state and take an intensive look at thirteen states both pre- and post-

2: Research on PRWORA: What Can Be Learned

PRWORA. Much of the Urban Institute project will center on a pre-PRWORA survey and several post-PRWORA national household sample surveys—the National Survey of America’s Families (NSAF)—with subsamples large enough to support precise estimates for each state in each of the thirteen states in which intensive administrative studies are being undertaken. NSAF also will oversample households at 200 percent or less of the poverty level, supporting detailed findings for low-income families. Additional surveys are planned for 2001 and, tentatively, for 2003.

- **Project on Devolution and Urban Change.** The Manpower Demonstration Research Corporation (MDRC) is undertaking the Project on Devolution and Urban Change, which consists of intensive studies of poor households in four major cities—Cleveland, Miami, Philadelphia, and Los Angeles. Using administrative data, the program participation and earnings of cohorts of poor households entering the Food Stamp or AFDC/TANF programs at given points in time will be tracked before and after TANF went into effect. In addition, it will include surveys of a sample of AFDC/TANF households, intensive studies of the TANF programs in the four cities, institutional analyses, and ethnographic studies within selected poor neighborhoods.
- **Child Impact Waiver experiments.** The U.S. Department of Health and Human Services (HHS) is augmenting five ongoing randomized waiver experiments with measures of effects on children in projects referred to as the Child Impact Waiver experiments. The five waiver experiments began before PRWORA and compare randomly selected families who have been receiving welfare benefits under the old AFDC rules (control-group families) with randomly selected families who received benefits under the waivers (experimental-group families). Because these extensions are among the best of the waiver experiments and will provide information on effects on children, they are considered a fourth major research project for the purposes of this paper. Most important, the waiver provisions being studied resemble quite closely the TANF provisions enacted by the states being studied. The five states and the contractors involved are Florida (MDRC), Minnesota (MDRC), Connecticut (MDRC), Iowa (Mathematica Policy Research), and Indiana (Abt Associates).

The question that most interests the policy community is: What have been the net effects of TANF (that is, effects uniquely attributable to TANF) on the employment and well-being of low-income households? The “gold standard” design for estimating net effects is the randomized experiment. Only the Child Impact Waiver studies are experiments; the other three research projects provide nonexperimental data that cannot support estimates of net effects that are as credible. Because they must rely on “before-and-after” research analyses, they will be unable to distinguish between the effects of welfare reform and those of other economic and policy changes occurring at the same time. The best use of the nonexperimental research projects is to provide descriptive information about what has happened to low-income households and families as they experience TANF.

2: Research on PRWORA: What Can Be Learned

Weaknesses in Research Design

Each research project suffers from inadequacies that diminish its usefulness to greater or lesser degrees, as follows:

- **Survey of Program Dynamics.** The SPD's most attractive feature is its promise to provide longitudinal data on national samples of households for five years before and after the enactment of PRWORA. However, the most serious problem facing the SPD is a low response rate, with just 50 percent of the original sample responding in 1998 and 1999. Especially worrisome is that higher percentages of low-income households stopped cooperating. Furthermore, it is likely that less than one-quarter of the sample will produce complete longitudinal data covering all or even most of the interviews over the ten-year period. The Bureau of the Census has begun efforts to reach and obtain the cooperation of the nonresponding households, hoping to raise the response rates to above 60 percent. Even if efforts to reduce future attrition are that successful, the rates will not be high enough to satisfy most researchers.
- **Assessing the New Federalism.** A central component of the Assessing the New Federalism project is the NSAF, a series of nationally representative surveys conducted in 1997 and 1999, with additional surveys planned for 2001 and, possibly, 2003, to permit comparisons between households before and after welfare reform. Response rates for the first NSAF survey were average for well-run national telephone surveys—65 percent for families with children and 62 percent for families without children, but “average” may not be good enough for surveys that are highly policy relevant.¹ NSAF staff have attempted to compensate for low response rates by weighting the data. The large samples for each of the thirteen states also make it possible to examine how families in states with different TANF plans have fared. However, the decline in welfare rolls has considerably reduced the number of welfare families in the survey taken after TANF went into effect. The small sample sizes within each state will restrict the ability of analysts to estimate the impact of welfare reform, especially subgroup differences at the level of individual states.
- **Project on Devolution and Urban Change.** MDRC's Project on Devolution and Urban Change focuses on welfare recipients in four major cities, combining administrative data from 1992–2002 with surveys of welfare mothers who were on the rolls in 1995 and a yet-to-be-determined year in the post-PRWORA era. Response rates for the first survey in 1997 were very good, averaging 79 percent across the four sites. The project should provide a rich—and textured—picture of the characteristics and circumstances of the most vulnerable families in the cities studied. However, the findings cannot be generalized to the broader welfare population nationally or even to other urban neighborhoods.

¹Editor's note: In their response to Rossi, the Urban Institute researchers argue that comparisons to other surveys should be based on an alternate “weighted” response rate that is slightly higher (about 70 percent).

2: Research on PRWORA: What Can Be Learned

- **Child Impact Waiver experiments.** HHS is augmenting five ongoing waiver randomized experiments, comparing families who have been receiving welfare benefits under the old AFDC rules with families who received benefits under the state's welfare reform plan. The importance of these continuing experiments is considerable. Because the experiments resemble each state's TANF program and are evaluated using a rigorous evaluation, they represent perhaps the best attempt at measuring the impact of welfare reform in selected states. Although the projects all rely on the universally preferred randomized experiments, generalizing findings to other states is clearly hazardous. It also is not clear whether the experimental conditions have been maintained with fidelity. Given the widespread publicity accompanying welfare reform, it is possible that control-group families may not have been subjected to AFDC rules and that such families did not fully realize that their welfare benefits were not subject to state TANF rules.

What Will and Will Not Be Learned from the PRWORA Research Projects?

Assuming successful completion of data collection for the four projects, what will be learned from them? There can be little doubt that once released, the descriptive findings will be met with intense interest. Public and partisan interest in how the poor are faring under PRWORA is strong. These studies will provide that desired information in great detail.

- **National-level descriptive findings.** Both the SPD and Assessing the New Federalism will provide findings concerning how American households were doing socioeconomically before and after PRWORA. For the nation as a whole, we will know whether earnings and employment of low-income households improved or declined. We also will know whether in certain respects the children in low-income households are better or worse off.
- **State-level descriptive findings.** Assessing the New Federalism will provide detailed information on thirteen states, and the Child Impact Waiver experiments will provide information on five states.
- **Local-level findings.** The Project on Devolution and Urban Change will provide detailed descriptive data on the welfare population of four large cities. The five Child Impact Waiver experiments will provide similar information in selected localities.

The prospects for information on the changes in the condition of the poor that can be credibly attributed to PRWORA are not as rich. The five Child Impact Waiver experiments will provide good impact estimates for the five TANF-like programs, but the extent to which those findings will be able to be generalized to the nation as a whole is limited.

As for the nonexperimental studies—SPD, NSAF, and Urban Change—their before-and-after designs generally will not support very credible impact assessments. Nevertheless, some attempts to estimate the net effects of PRWORA generally, or TANF specifically, may turn out to be

2: Research on PRWORA: What Can Be Learned

more than merely suggestive, especially when findings are strong, consistent, and robust. For example, if analysis using NSAF data were to find that states with more generous earnings disregards had relatively higher caseloads, holding other interstate differences constant, and that such effects were muted when combined with a strict time limit or work requirement, then the findings could be regarded as supporting a causal inference, if they held up under different specifications and were confirmed in comparable analyses in the SPD and UC data sets, as well as in the Child Impact Waiver experiments. Such a convergence of findings across data sets, however, is not likely to occur often.

Other problems hamper research on PRWORA. First, devolution means that each state can design a different version of TANF. As a result, PRWORA cannot be evaluated as a national program—only state programs can be evaluated. Second, state programs can be expected to change over time as states change them in reaction to experiences. Third, larger forces compete with the programs. If labor market conditions for entry-level jobs change over time, those changes may affect the employment of welfare clients more than TANF. Fourth, some of the effects of PRWORA can be expected to be manifest quickly but others may take a number of years to show.

Welfare Reform Research in the Future

In the near term, it is important to support some of the ongoing research projects by providing additional resources to strengthen their contributions. In particular, SPD will not be very useful unless response rates can be materially improved. Accordingly, it is important that the funds the Bureau of the Census needs in order to raise the response rates to SPD be appropriated. It is heartening that some funds are already at hand, but the bulk of the funding is not yet forthcoming. If SPD can be materially improved, it will provide extremely useful information on the changes accompanying welfare reform.

The effort to raise response rates should be monitored carefully: If at some point it becomes clear that response rates will not be materially improved, then the effort ought to be discontinued and the unexpended funds put to some better purpose (see below). In addition, serious consideration ought to be given to actions ranging from releasing the data sets with strong warnings about their limitations to suppressing their release entirely.

The prospects for improving NSAF through investment of additional resources do not appear to be good. The Urban Institute and Westat have done as much as possible to compensate for NSAF's weaknesses. The low response rate to NSAF-I is troubling and is irremediable except through careful weighting. However successful the weighting scheme may be—and even if NSAF-II has achieved a more acceptable response rate—two surveys, before-and-after, are quite a weak design for estimating net effects. The major value of the Urban Institute studies will come from the detailed descriptive data for the thirteen states. It will be difficult to capture the diversity of state policies in a statistical model, given the many variations chosen by states and the frequency of their change.

2: Research on PRWORA: What Can Be Learned

MDRC's Project on Devolution and Urban Change is not far enough along to make any judgment concerning its prospects. Especially critical will be its ability to analyze cohort experiences using administrative data. Assuming successful statistical modeling of quality data, Urban Change will provide good estimates of effects within four important localities, supplemented by qualitative data on four local welfare systems. It is clearly too early to judge whether any steps can be taken in the short term to improve Urban Change or, indeed, whether improvements will be needed.

The ongoing waiver experiments could prove to be valuable if strong efforts are made to ensure the fidelity of control-group conditions in each experiment. The danger is great that the control groups will be treated inappropriately and that the members of the control groups will not understand that they are not subject to state TANF rules. Maintaining the integrity of the control groups means not only more effort on the part of the contractors but also the possibility that additional funds need to be given to support training of agency personnel and to provide for more frequent reminders to control-group members of the special rules governing their welfare benefits. Unfortunately, given that the experiments have been underway for nearly four years since the start of TANF, it may now be too late to bolster their integrity.

It is also quite clear that when the four projects have been completed and analyzed, their findings will leave many critical questions unanswered. Almost certainly, PRWORA will be shown to be successful in meeting some of its goals in some of the states and failing to meet other goals in others. Questions will be raised about the effectiveness of time limits, family caps, income disregards, and other elements of the reform bundle. To answer those questions, further research will be needed. What form should such research take?

Perhaps the best strategy over the next decade or so would be to authorize and fund randomized experiments testing variations on the administrative and policy bundles. Those studies could be accomplished through state initiatives, but they are not likely to happen without federal funding. For example, questions about the effectiveness of family caps in reducing fertility might best be answered by conducting randomized experiments in which the experimental group is subject to family caps and the control group is under a no-family-caps condition (or vice versa). Randomized experiments can also be designed to observe the effects of varying the generosity of income disregards. Factorial experiments might be started to test the effects of various combinations of provisions that make up administrative and policy bundles.

This research strategy should lead to the accumulation of knowledge about how best to design welfare to achieve the dual objective of providing a safety net for the poor and facilitating entry into employment and higher income. A more expanded version of the current strategy of HHS's Administration for Children and Families (ACF) is proposed here. The current ACF experiments are not designed to unbundle TANF as much as to test proposed additions to the

2: Research on PRWORA: What Can Be Learned

bundle. To understand how the bundles work, it is necessary to design experiments that vary such critical elements as earnings disregards, family caps, and time limits.

Some progress has been made toward implementing this strategy: ACF has funded three experiments and has issued an RFP for a fourth to be funded in 2000. All the experiments are designed to test measures aimed at improving TANF. For example, an experiment in Virginia will test the effectiveness of postemployment services in helping TANF clients retain their newly obtained employment. An evaluation planned in the future will involve four to ten states in MDRC-run experiments on measures aimed at employment retention and advancement in employment. The new experiments are patterned after the waiver strategy followed in the last decade or so of AFDC.

Studies also are needed to provide detailed descriptions of how the poor will fare under the welfare policy changes instigated by PRWORA and under whatever other policy changes occur. We should be planning now how best to collect the data that will support an empirically based understanding of what is happening to the poor and what policy changes are likely to improve their condition.

When AFDC was a more or less uniform national program, national surveys such as the Current Population Survey (CPS) or the SIPP may have served the purpose of monitoring the well-being of the poor. However, as discussed earlier, devolution has meant that state-level rather than national-level data are needed. An obvious move would be to enlarge sample sizes of existing ongoing national surveys to provide adequate state sample sizes. NSAF provides a good example in its selection of a small sample of critical states, an approach that the national surveys, including SPD, might want to emulate.

It is likely that in the end, SPD will not be very useful. Hence, serious consideration ought to be given to bolstering other ongoing large-scale surveys. In particular, it would be useful to augment SIPP and CPS by enlarging their sample sizes, especially bolstering their coverage of poor families. Ideally, I would like to see the sample sizes in at least the largest states increased enough to support state estimates.

Up to this point, the CPS has provided good monitoring data on the condition of the poor for the nation as a whole. Expanding the CPS sample to provide detailed data on a sample of states—and expanding CPS variables to include more information on how families with children are faring would be extremely useful. Additional efforts also should be made to address the problem of underreporting of welfare receipt. A parallel expansion of SIPP to conduct annual panel studies in a sample of states, especially in the ten to fifteen states that contain most of the poor, would be able to provide information on post-PRWORA changes in some detail. I recommend that the National Research Council of the National Academy of Sciences or a similar body examine the suitability of using SIPP for this purpose, paying special attention to attrition and nonresponse and their impact upon obtaining valid and reliable analyses.

2: Research on PRWORA: What Can Be Learned

Finally, a serious issue is how to promote responsible analyses of these data sets. Neither the research nor the policy communities will be content with only descriptive analyses. If Wisconsin poor families are better off (or worse off) in 2001 than they were in 1997 but California poor families show an opposite pattern, then some analysts certainly will try to discern whether the differences between the two states' versions of TANF are the source of the difference. To some extent, we can expect that competition among analysts will provide constructive criticism. In any event, those who release public data sets should warn potential users about the limitations of their data as well as provide full and detailed documentation about the data sets.

RESEARCH ON PRWORA

After decades of relative stability in structure, public welfare began to change in radical ways in the 1980s. Under encouragement from the Bush and Clinton administrations, state welfare departments were urged to apply for waivers permitting them to depart from federal regulations to try out new ways of providing support to the poor under Aid to Families with Dependent Children. Most states applied for and were successful in receiving waivers. When the Personal Responsibility and Work Opportunity Reconciliation Act was enacted in August 1996, many of the states' changes—and additional modifications—were institutionalized.

In the widest sense, the welfare reform receiving so much attention today consists of the cumulative changes made over the past ten to fifteen years. "Welfare as we knew it" is really welfare as it was in the late 1980s. By the time PRWORA was enacted, the "old" welfare in most states had been transformed, in some instances radically so.

An important change in welfare also was underway by the time PRWORA was enacted. Average monthly enrollment in AFDC began to decline in 1994 and continued to decline after PRWORA. Some of the decline can be attributed to economic prosperity and the accompanying high employment rates, but as the decline approached 50 percent by mid-1999 (compared with 1994), it also became apparent that fewer families were applying for welfare. It appears that in the early years, the decline went largely unnoticed in either policy or welfare research quarters. Perhaps PRWORA accelerated the downward trend in enrollment, but perhaps not. In any event, this long-term trend would become an important aspect of the social changes accompanying PRWORA and, as will be discussed later, presents a vexing problem for existing welfare reform research.

The major provisions enacted under PRWORA were:

- The abolition of AFDC as an entitlement program, to be replaced by TANF, a time-limited welfare program with federal support to each state limited by historical funding patterns.
- Strong emphasis on moving TANF recipients off the rolls and into employment.

2: Research on PRWORA: What Can Be Learned

- Discretionary powers given to the states to design their own versions of TANF. Significant but less drastic changes were made in other welfare programs such as the Food Stamp Program, Medicaid, and Supplemental Security Income.

Within broad limits the states may design their own versions of TANF. Some states have elected to continue policies they put in place under AFDC waivers received prior to the passage of PRWORA. Other states have adopted shorter time limits than required for TANF. All states have changed, although the changes in each state differ. We can safely anticipate that more changes will occur over the coming years as states try new policies, find some wanting, and move on to implement modifications.

Although federal regulations can be expected to produce some degree of uniformity across states, it is likely that no two states will develop exactly the same set of provisions in their welfare programs over the next few years. Furthermore, it also is likely that many states will modify and change their programs in response to implementation problems or pressures from interested constituencies. It is also possible that federal TANF program requirements might change: For example, if time limits were found to produce extreme hardships for some families, Congress might very well modify those provisions or allow the states to do so. In addition, some states may establish state-financed programs designed to mitigate perceived hardships for some groups of poor families. Indeed, some states assumed the costs of providing food stamps to legal immigrants when the initial provisions of PRWORA declared food stamp eligibility restricted to citizens.²

The evaluations of the effects of some of the waivers put in place over the past decade can provide at least some hints about the effects to be expected from some features of the state TANF programs. Among the more promising waiver experiences are those that were put in place accompanied by randomized experiments intended to evaluate their effects and which were continued after PRWORA took effect. More than a dozen states have continuing AFDC waiver experiments. Strong inferences about TANF effects can be made for states in which the waiver provisions being tested closely match provisions adopted under TANF. In other continuing waiver experiments, mismatches make convincing inferences about TANF effects more difficult. Although the best of the waiver experiments will provide useful information on TANF effects, the findings will be limited to a handful of states and to comparisons between TANF and AFDC. In any event, a need will remain for more information on TANF in all its various manifestations throughout the country.

At the point of enactment, both the nature and the consequences of the changes enacted under PRWORA were largely unknown (and remain so). Perhaps the changes would lead to improvements, but perhaps not; the changes might very well produce some unanticipated effects. New research efforts were needed to find out what shape welfare reform would take in each of the states, the effects of the new welfare systems on public and private agencies and, most of all, the

²This provision was modified in 1998 to allow most naturalized citizens to qualify for the Food Stamp Program.

2: Research on PRWORA: What Can Be Learned

effects on poor households and the children within them.³ The four research projects reviewed in this paper were designed to provide the needed empirical answers. For reasons discussed in some detail in the next section, none of the projects were easy to design, and none will be easy to carry out and analyze. Furthermore, any research conducted on the scale necessary to be relevant to all states or even a few key states will be expensive and take years to complete.

For all the research projects, the evaluation of PRWORA is primarily the evaluation of the additional changes embodied in that legislation and in the welfare plans designed by the states. It is not the evaluation of welfare reform represented by all the changes that began taking place before 1996. The decline in welfare rolls beginning in 1994 illustrates that the pre-PRWORA welfare changes may have had effects that precede those of PRWORA. In addition, the research projects are concerned mainly with the TANF provisions of the PRWORA legislation. Accordingly, the studies reviewed here are mainly concerned with TANF changes imposed on top of an already changed welfare system.

Of the four major research projects that this paper reviews, three were designed specifically to study the effects of PRWORA/TANF—we will call them “prospective” studies. The fourth project consists of extensions to five waiver experiments that were underway before PRWORA and are being modified to be relevant to TANF—we will treat them as a group, calling them “ongoing” studies.

The three major prospective research efforts currently underway⁴ are designed on scales large enough to promise to generate findings relevant to significantly important places or states:

³ Existing ongoing research efforts were judged to be deficient for these purposes. The Current Population Survey (CPS) collects no information on child well-being and was not designed to provide sufficiently large samples for more than a handful of states. Both the National Health Interview Survey and the Survey of Income and Program Participation have inadequate sample sizes, especially for individual states.

⁴ Several additional efforts not covered in this paper are underway:

- The Rockefeller Institute (SUNY-Albany) has started the states Capacity Study project, a study of the administration of the new welfare programs in several states. Because this effort is not directly concerned with measuring the effects of such changes on poor families, it is not reviewed in this paper.
- A group of prominent social scientists (Ronald Angel, Linda Burton, P. Lindsay Chase-Lansdale, Andrew Cherlin, Robert Moffitt, and William J. Wilson) have begun a longitudinal study of poor families based on samples taken within Boston, Chicago, and San Antonio, supplemented by ethnographic studies in poor neighborhoods within each of those cities and child development studies of welfare children. This collaborative effort was funded in 1998, and fieldwork has been underway since early 1999. In addition, an inventory and description of ten major surveys relevant to family well-being post-PRWORA can be found in Brick (1999).

2: Research on PRWORA: What Can Be Learned

- The Bureau of the Census is extending the Survey of Income and Program Participation (SIPP) to create the Survey of Program Dynamics (SPD), a longitudinal survey consisting of repeated interviews of a national sample of households from several years before PRWORA to several years after enactment. SPD oversamples low-income households.
- The Urban Institute's extensive study, *Assessing the New Federalism*, is examining the devolution shifting many welfare programs from the federal to the state level. Administrative studies will describe the changes in health and welfare systems in each state and take an intensive look at thirteen states both pre- and post-PRWORA.⁵ Much of the Urban Institute project will center on a pre-PRWORA survey and several post-PRWORA national household sample surveys—the National Survey of America's Families (NSAF)—with subsamples large enough to support precise estimates for each state in each of the thirteen states in which intensive administrative studies are being undertaken. NSAF also will oversample households at 200 percent or less of the poverty level, supporting detailed findings for low-income families. Additional surveys are planned for 2001 and, tentatively, for 2003.
- The Manpower Demonstration Research Corporation is undertaking the Project on Devolution and Urban Change, which consists of intensive studies of poor households in four major cities: Cleveland, Miami, Philadelphia, and Los Angeles. Using administrative data, the program participation and earnings of cohorts of poor households entering the Food Stamp or AFDC/TANF programs at given points in time will be tracked before and after TANF goes into effect. In addition, the study will include surveys of a sample of AFDC/TANF households, intensive studies of the TANF programs in the four cities, and ethnographic studies within selected poor neighborhoods. The first survey was completed in 1999.
- The U.S. Department of Health and Human Services (HHS) is augmenting several ongoing waiver experiments, adding measures of effects on children.⁶ The five studies, known as the Child Impact Waiver experiments, compare randomly selected control-group families who have been receiving welfare benefits under the old AFDC rules with randomly selected experimental-group families who received benefits under the waivers. Because these extensions are among the best of the waiver experiments and will provide information on effects on children, they are considered a fourth major research project for the purposes of

⁵ Although NSAF interviewing took place in 1997, the survey collected information about 1996, covering the period before most PRWORA provisions went into effect.

⁶ In addition to the five ongoing waiver experiments discussed in this paper, HHS is funding the continuation of experiments in Vermont, Texas, and Arizona. All eight waiver experiments were chosen for continuation because waiver provisions closely resembled their states' subsequent TANF plans. The five ongoing waiver experiments chosen to be discussed in this paper were selected because they each added child impact measures, bringing them more into line with the prospective studies.

2: Research on PRWORA: What Can Be Learned

this paper. Most important, the waiver provisions being studied resemble quite closely the TANF provisions enacted by the states in question. The five states and the contractors involved are Florida (MDRC), Minnesota (MDRC), Connecticut (MDRC), Iowa (Mathematica Policy Research), and Indiana (Abt Associates).

The central purpose of this paper is to assess the prospects of the four research projects for providing credible findings on how the TANF programs will affect low-income families. Because none of the prospective research projects have gone beyond the first steps in data collection, the paper focuses on the ability of their research designs to support credible findings concerning the state welfare programs and their effects. For research programs that are already engaged in data collection, the paper examines how well the designs have been implemented. Of course, none of the evaluations have proceeded far enough to produce any findings relevant to TANF, and the research designs of the prospective studies are not yet in their final forms.

The four research projects each have several objectives, some of which are unique to each study. As far as evaluating welfare reform is concerned, there are three central objectives:

- **Describing how public welfare changed under PRWORA.** MDRC, the Urban Institute, and the waiver experiment studies share the objective of describing how public welfare changed in each of the sites studied, whereas the Census Bureau's SPD does not.
- **Describing how low-income families fared before and after welfare reform.** In this area, the studies will give close attention to levels of employment, earnings, TANF reciprocity, family functioning, and child well-being. All four projects share this objective, although MDRC primarily will use program administrative data supplemented by two cross-sectional household surveys, SPD will be a panel study, and NSAF will rely on three (possibly four) cross-sectional household surveys. The Child Impact Waiver experiments deal primarily with families who have been on welfare and not the entire population of low-income families.
- **Estimating the net effects of welfare reform.** Only the Child Impact and MDRC projects currently intend to estimate the net effects of TANF or other changes enacted by PRWORA. The Bureau of the Census is not planning to analyze the SPD data set in detail but will make it available for public use. The Urban Institute is not quite settled about analysis plans but will make its data set available for public use. Because the three latter data sets will be made available for public use, all likely will be used by a variety of analysts to estimate net effects.

Findings bearing on all three topics undoubtedly will be met with great interest on the part of everyone interested in welfare reform. The greatest concern is likely to be with estimates of net impact. Surely changes will occur in how clients fare under the new welfare system, but the central question is whether those changes can be credibly attributed to the changed welfare system or to other forces or events.

2: Research on PRWORA: What Can Be Learned

In addition, the first and second descriptive tasks, although demanding great care and skill, can be accomplished in fairly straightforward ways. System changes can be discerned in documentary materials and supplemented by interviews with key welfare officials—and, perhaps, with rank-and-file welfare workers. The methodology lying behind sample surveys is well known; although household surveys are not easy to conduct, given sufficient resources, excellent surveys can be conducted. In contrast, estimating the net effects of a full-coverage social program is much more difficult, for reasons discussed in the next section.

Why Estimating the Net Effects of PRWORA Will Be Difficult

This section describes the difficulties that will beset attempts to estimate the effects of PRWORA. The principals of the four prospective projects are very much aware of the problems and discuss them at length in project documents. Because the problems facing the ongoing waiver experiments differ from those facing the prospective studies, they are discussed in a separate section.

Urban Change, Survey of Program Dynamics, and Assessing the New Federalism. The net effects of a program are estimated by measuring relevant outcomes among program participants and comparing those outcomes with what would have been measured had the subjects not experienced the program. In a literal sense, such comparisons are impossible. The best approximation is achieved by conducting randomized experiments, in which subjects are randomly assigned to different groups that either experience the program or serve as controls from which the program is withheld. The conditions to which the control groups are subjected are known as *counterfactuals*. Because the assignment to experimental and control groups is random, the two groups differ only by chance. Hence, measures of outcome that differ between the two groups to an extent greater than what can be expected from chance can be reasonably attributed to the effects of the program. Program net effects are always estimated comparatively. The comparison is between the program being evaluated and the “program” represented by the conditions prevailing in the control group.

In principle, randomized experiments can be used to estimate the effects of PRWORA or TANF, although it would be extremely difficult to carry them out for a variety of reasons. First, it is not clear what counterfactual conditions should apply to control groups. Full coverage (or saturated) programs, such as PRWORA or TANF, present special difficulties because such programs cover all persons or households who are eligible. A control group cannot be formed without subjecting households assigned to control conditions to some policy-relevant alternatives.

The quintessential evaluation question is whether a given program is effective compared with an alternative program. In the usual randomized experiment, outcomes in the experimental group who have experienced a program are compared with those in a control group who have experienced some other treatment, usually the status quo. For example, in the welfare-to-work

2: Research on PRWORA: What Can Be Learned

experiments conducted under AFDC waivers, the experimental groups each experienced a new program designed to facilitate the movement of welfare clients into paid employment, whereas the control groups experienced the conventional state welfare program, usually AFDC. The policy issues were whether and how the changes embodied in the waiver provisions differed from AFDC in their effectiveness.

It is not clear how the control conditions should be defined in a randomized experiment involving TANF (Hollister 1997): Should the control condition be a pre-1996 AFDC program? If so, then the experimental findings compare TANF with a condition that most observers thought was ineffective and to which few would want to return. A study showing that compared with AFDC, TANF did not produce a significantly greater movement into employment would be unlikely to lead to a clamor for a return to AFDC. Rather, it would lead to proposals for modifying TANF to strengthen its effects on employment, for which the research findings would provide little guidance. Strong arguments for using AFDC as the counterfactual control-group condition also can be made, however. Critics of TANF have claimed that poor families will become worse off, suggesting that AFDC is an appropriate comparison group for TANF studies.

The issue of how control conditions should be defined need not be settled entirely one way or the other. Comparisons with AFDC are needed to settle the issue of whether TANF has deleterious effects on the poor. We also need to test how TANF can be improved. The latter question suggests that a randomized experiment strategy would be to test the effectiveness of alternatives to TANF, defining the control condition as the TANF status quo.⁷ For example, some consensus may exist among policymakers that time limits of some kind are going to be a feature of any alternative to TANF, but it may not be clear whether longer or shorter time limits will be more effective. A sensible approach might be to define the experiment as testing the relative effects of longer and shorter time limits than those currently specified. Note that in this approach, the experimental intervention is variation in the size of time limits, and the control condition is the TANF status quo. The research question then becomes, “Can we improve on TANF?” No experiments using this strategy currently are underway or planned.⁸

⁷ A demonstration quasi-experiment called Jobs-Plus is an example of such research (Riccio 1998). Jobs-Plus, conducted by MDRC and financed by the U.S. Department of Housing and Urban Development and the Rockefeller Foundation, is based in eight public housing sites throughout the country and is testing the ability of a combination of community organization, job training and job search activities, and enhanced incentives for working to move welfare families living in public housing into employment. The housing projects for the demonstration were selected randomly from among a set of three roughly equivalent projects in each of eight cities, with two projects designated as controls. The main variations on TANF consist of community organization efforts and enhanced incentives. Note that Jobs-Plus is primarily a variation on TANF for public housing communities and is not intended to be broadly applicable to TANF. Of course, if Jobs-Plus proves effective, that finding might suggest that community strategies should be considered in TANF.

⁸ It is my understanding that the Administration for Children and Families of HHS has provided grants to thirteen states to plan, design, and evaluate variations in TANF provisions for improving job retention

2: Research on PRWORA: What Can Be Learned

In all three of the prospective research projects this paper reviews, the main contrasts that will be available will be between how households were faring before 1997 under AFDC and their condition after living under TANF. Of course, the studies also can be used as cross-sections. For example, NSAF data on the thirteen states can be used to estimate the effects of state TANF plan differences on the well-being of poor households. Or, children in TANF families can be compared with children in comparable non-TANF families. Such analyses can hardly ever be credible and are certainly far inferior to even before-and-after analyses. Accordingly, they are not discussed in detail in this paper. All can be analyzed with nonexperimental research designs, using as nonrandomized controls or as comparisons the experiences of poor households pre-TANF (essentially AFDC in 1996 or earlier).

These before-and-after research designs are perhaps the best designs possible under the conditions existing when the research programs were formulated—the projects had to get rapidly underway before PRWORA was enacted but when it was clear that some sort of welfare reform was inevitable. Consequently, the design options were limited. Nevertheless, the findings resulting from such designs are widely viewed by evaluators as subject to many credibility problems. The most serious problems are twofold.

First, the client population may change as a consequence of TANF. Indeed, the welfare rolls began to decline in 1994, several years before the enactment of PRWORA, hinting that the composition of welfare clientele under TANF might be different in important respects compared with AFDC clientele. At least some of the decline in caseloads appears to be because some families who might be eligible for benefits have decided not to apply for them. Such “entry effects” cannot easily be studied or even detected in the MDRC and Urban Institute projects, especially the conditions of those families who were “deterred” from applying for welfare.

Second, before-and-after designs are not able to distinguish convincingly the effects of interventions from other changes occurring at the same time. For example, the studies may detect a significant increase in employment of single mothers after TANF, but that increase may be a result of increased employment opportunities resulting from a tighter labor market. Changes in other programs, such the Earned Income Tax Credit, or changes in the federal minimum wage also may affect employment.

Some of the difficulties in estimating TANF effects lie in the nature of the reforms themselves. Within quite broad limits, each state can develop its own version of welfare reform. For example, some states have quite generous income disregards, whereas others are less generous; some have quite short time limits on eligibility, but other states permit longer periods of eligibility.

and advancement for TANF recipients.

2: Research on PRWORA: What Can Be Learned

Some have imposed family caps, which involve not increasing income maintenance payments when children are born after initial eligibility. And so on.⁹

In effect, TANF cannot be evaluated as a national program; only state TANF programs can be evaluated. In addition, one can expect within-state variation in implementation, especially in such states as California or New York, where welfare is administered by local political jurisdictions (for example, counties), with some local discretion allowed in design. In most other states, TANF plans permit considerable discretion to “street-level bureaucrats” in implementation, potentially leading to within-state heterogeneity.

Within each state, TANF is designed as a package of interventions, typically consisting of several provisions, some of which can be expected to have quite opposite effects. For example, generous earned income disregards may make employment quite attractive, leading to higher household incomes for some households, but sanctions imposed for noncompliance with work provisions may serve to reduce income for other households. Accordingly, if a research project judges a state’s program to be effective, it will be quite impossible to attribute those effects to any specific provision in the “bundle” that constitutes that state’s version of TANF because only the bundle can be evaluated.¹⁰ Furthermore, the effects of some provisions may be counterbalanced by the opposite effects of other provisions to the extent that the bundle may appear to have had no effects. The consequence is that it will be difficult, if not impossible, to discern which elements of each state’s TANF bundle are contributing (and how much) to success or failure.

Not only the state bundles but also the names given them vary. Although “Temporary Assistance for Needy Families” is the title given in the federal legislation to the successor to AFDC, some of the states have given different titles to their versions. This diversity produces difficulties for national surveys. It will be hard to write questions about TANF program participation in national surveys so that they are understood properly by respondents.¹¹

⁹ The diversity that TANF state plans have developed are described in detail in a recent Urban Institute report summarizing state plans as of October 1977 (Gallagher et al. 1998).

¹⁰ In principle, it is possible to study the separate effects of provisions making up a bundle. Factorial experiments using several experimental groups, each consisting of unique combinations of programs, would provide estimates of the relative effectiveness of the separate programs and combinations of programs that make up a bundle. Factorial experiments have been used in the two major income maintenance studies (Rossi and Lyall 1974; Robins, et al.1980) as well as in the Housing Allowance experiment (Friedman and Weinberg 1980); one of the waiver experiments, described below, has a factorial design.

¹¹This difficulty may be one of the main reasons that the CPS’s estimates of TANF participation increasingly has departed in recent years from counts obtained from administrative data.

2: Research on PRWORA: What Can Be Learned

TANF will take some years to settle into a relatively steady state. HHS published final regulations in 1999; accordingly, each state's plans did not begin to take final form until that year.¹² Implementation failures and modifications undoubtedly will occur in some states' TANF programs. For example, several states have experienced difficulties in training frontline personnel, who under AFDC primarily determined welfare eligibility, to take on the task of facilitating client movement to employment (General Accounting Office 1998b.) As a result, the versions of TANF experienced in the first few years will change after a few years of trial and error. All three research projects are designed to measure what happens in the first few years of PRWORA and may therefore be measuring early versions of those programs rather than what will appear after the programs have settled down. Of course, determining how long such a period of adjustment will take is problematic.

It is also possible that some or all of the state PRWORA plans may never reach a steady state but will change continually over time. In addition, Congress may alter provisions in PRWORA, forcing changes on the states. Although evaluation of a frequently changing program can be undertaken, interpreting findings will be quite difficult.

Even if TANF functions as its authors expected, its fate depends heavily on forces outside the influence of the program. In particular, moving TANF clients into jobs cannot succeed unless jobs are available in the economy for which clients can be considered and unless the child care industry provides enough slots to meet the needs for child care. To be sure, the states might provide public-sector positions if the labor market does not offer private-sector employment, but arranging for and financing public-sector jobs will be difficult and expensive. Much more problematic is whether the child care industry can provide slots for the increased demand at prices that are affordable under state subsidies as well as meet quality standards. Unfortunately, none of the research projects plan to collect data directly relevant to those issues. For example, the sample surveys will collect data on respondents' child care arrangements, but no studies of child care markets are planned.¹³

Finally, issues of what should be regarded as criteria of success and when measurement should be undertaken remain. TANF has multiple goals as well as the potential to produce unanticipated and unwanted effects. The key outcome of greatest interest concerns the extent to which welfare clients will move off the rolls, obtain employment, and thereby be better off. Because clients are expected to be better off when employed, researchers are interested in the earnings of those who move into employment. Because some analysts have predicted that children will be worse off under PRWORA, they also have considerable interest in measures of child well-being. Accordingly, it is appropriate to ask whether each of the projected studies has planned to collect data relevant to the important outcomes that PRWORA is intended to affect.

¹² Nevertheless, many states wrote their versions without benefit of the federal regulations and had to change them to conform to the final federal regulations.

¹³ HHS is funding research to be conducted by Abt Associates on low-income child care markets; the planned survey may provide the needed information.

2: Research on PRWORA: What Can Be Learned

The outcome measures used are restricted to those that can be obtained by the data collection modes used. For example, the quality of day care used by the households studied cannot be measured directly, only through the reports of household respondents.¹⁴ Nor is it possible to measure directly other important outcomes, such as the quality of parent–child relationships or the cognitive abilities of children.

Some effects can be expected to appear relatively quickly, but others can be expected to occur after several years or longer. If TANF increases employment and earnings, such effects should appear within the first two years or so, whereas effects on the developmental trajectories of children may take a much longer period of time to detect. Steadiness of employment also requires observations over longer periods of time, whereas determining whether clients have found employment might require only a single interview. No one can expect that initial earnings of new entrants in the labor force will be high: It will take time to observe what shape earnings will take over time.

Ongoing waiver experiments. Relating the five ongoing waiver experiments to TANF presents somewhat different problems. Because each is a randomized experiment, the studies avoid all the difficulties stemming from nonrandomized designs, as described above. The waiver studies can produce unbiased estimates of the net effects of the waiver provisions involved in each case compared with the prewaiver AFDC programs in place in that state. The families in the experimental and control groups experience the same historical events and trends, so results will be independent of, say, trends in the local economy and labor market.

Using the waiver experiment findings to judge the effects of TANF, however, is problematic in several ways. First, although the waiver provisions being tested are fairly close to the TANF programs enacted in the five states involved, some differences exist, especially in time limits. Accordingly, ascertaining the implications of waiver experiment findings for TANF must be judgment calls, advanced through argumentation and clearly subject to dispute. (See the appendix for comparisons between TANF state plans and experimental treatments.)

Second, the waiver experiments are plagued by some of the same problems affecting the prospective studies. In particular, four of the plans are bundles of provisions that cannot be separately evaluated, the exception being the Minnesota experiment.¹⁵ The experiments also are studies of welfare reform in its earliest years.

¹⁴ Little empirically based knowledge exists on how to measure the quality of day care.

¹⁵ In Minnesota, the waiver provisions concerning mandatory services and the financial incentive are tested separately and in combination, making it possible to estimate the effects of each provision separately and in combination.

2: Research on PRWORA: What Can Be Learned

Third, the subjects enrolled in the waiver experiments will become increasingly unrepresentative of all welfare families over time. For example, the members of the households all will be five years older at the end of a five-year-long experiment—and likely to be older than current welfare clients. Any changes in the characteristics of welfare clients enrolled after PRWORA will not be detected. Accordingly, entry effects cannot be studied.

Fourth, maintaining the integrity of the waiver experiments will be difficult. In particular, it is essential that the control-group families be subject to AFDC rules throughout the time periods of the experiments. It also is essential that they know that they are under such rules. PRWORA and TANF have received so much general publicity in the mass media that control-group families may not know that time limits, family caps, and other TANF rules do not apply to them. In addition, control-group families are typically a small proportion of the total welfare caseload, and welfare personnel mistakenly may apply the wrong rules to such families. That this is a very real danger can be seen in the failure to maintain the fidelity of treatments for both the control group and the experimental group in the New Jersey waiver experiment evaluating that state's Family Development Program (Camasso et al. 1998a). A central feature of the New Jersey program was a family cap, under which family benefits were not increased when children conceived after enrollment were born to mothers enrolled in the plan. Despite the fact that the control group was not under the family cap rule, it was found that the benefits of more than a score of control-group families had not been increased after the birth of children. In addition, a survey of control-group families found that few knew that the family cap rules did not apply to them. Given these findings, it was clear that the experimental conditions had been violated to the extent that the resulting data can hardly be regarded as valid.¹⁶

The research organizations running the Child Impact Waiver experiments are fully aware of the need to be vigilant in maintaining the integrity of the experimental and control groups. Some have placed safeguards in the management information systems software, which block inappropriate actions with regard to the participating families. All are determined to provide reminders to participating families about the relevant rules to which they are subject. Perhaps these measures will succeed in maintaining fidelity in treatments—but perhaps not.

Issues of Research Design

Clearly, identifying research design problems is easier than devising effective countermeasures. Each research project has been designed to address the difficulties each expects to encounter. Those efforts and their likely successes or failures are described in some detail in the next sections.

¹⁶ The Family Development Program also was assessed using administrative data on all welfare clients before and after the onset of the program (Camasso et al. 1998b) .

2: Research on PRWORA: What Can Be Learned

Urban Institute: Assessing the New Federalism. Of the three welfare reform projects, the Urban Institute's *Assessing the New Federalism* is the most extensive.¹⁷ About \$66 million in grants from various private foundations constitute the initial funding. *Assessing the New Federalism* consists of three major research activities:¹⁸

- **State database.** The Urban Institute has constructed and made available on its web site a large database containing hundreds of data items relevant to welfare reform for each of the fifty states and the District of Columbia. The data items range widely, from state demographic characteristics and vital statistics to AFDC and General Assistance caseload data. Many variables are disaggregated and tabulated separately for important population subgroups. Some of the series enable tracking changes over time, although usually they do not cover more than the previous ten years. The Urban Institute plans to add to the database as new measures become available. The state database is designed to serve as an information source to state and local policy analysts and welfare agencies as well as social scientists.
- **State administrative case studies.** Intensive case studies have been undertaken of the welfare and health systems of thirteen states (Alabama, California, Colorado, Florida, Massachusetts, Michigan, Minnesota, Mississippi, New Jersey, New York, Texas, Washington, and Wisconsin), which contain more than half of the U.S. population. The case studies will describe each state's welfare system before the full implementation of PRWORA and again in 1998 or 1999, after PRWORA changes were made. The case studies are based on state documents and legislation as well as interviews with key state administrative personnel. Reports describing each state's pre-PRWORA welfare and health systems on the basis of research conducted in 1996 and 1997 are being released as they are completed. The post-PRWORA case studies have been started and will serve as the bases for additional reports.¹⁹
- **The National Survey of America's Families.** NSAF is planned as a set of national surveys of American households. The first, NSAF-I, was conducted in 1997 after the enactment of PRWORA, but before full implementation. Many of the questions in the interview asked about conditions and events experienced in 1996, the year before the enactment of PRWORA. It was followed in 1999 by a second survey, NSAF-II. The two NSAF surveys will constitute a before-and-after research design enabling comparisons between household

¹⁷ This section is based on *Methodology Reports* 1–8 (Urban Institute 1999a-f) and on interviews and correspondence with Fritz Scheuren, Anna Kondratas and Genevieve Kenney. In addition, I attended, upon invitation, a February 1998 meeting of the project advisory group, at which the progress of the project was discussed.

¹⁸ A fourth major activity consists of dissemination of findings to the national and state policy communities.

¹⁹ A compilation of TANF provisions for all fifty states as of October 1997 has been issued (Gallagher et al. 1998).

2: Research on PRWORA: What Can Be Learned

measures from two cross-sectional surveys. NSAF-III is currently being planned for 2001, and NSAF-IV is tentatively planned for 2003. NSAF-III and NSAF-IV (if undertaken) will provide data on conditions experienced as PRWORA and TANF mature.

NSAF-I and NSAF-II are complex household surveys. Although their findings can be projected to the entire United States, they also provide large enough samples within each of the thirteen states chosen for the administrative case studies to make possible findings concerning each of those states. (More details on the surveys are given below.)

The three subprojects are intended to provide a comprehensive description of the changes in health and welfare policies resulting from devolution. The resulting data sets will support analytic studies of such important issues as whether the states have engaged in a “race to the bottom” by lowering benefits to the poor. Both the state database and the state case studies will provide valuable descriptions of the fifty states and the state administrative changes accompanying PRWORA. Although data on the central evaluation questions concerning the impact of PRWORA on low-income households can come from the two NSAF surveys, the administrative data sets will play important roles in explaining interstate differences in the NSAF findings. This review will concentrate primarily on describing and assessing NSAF.

NSAF-I and NSAF-II are quite similar in design. Each consists of two complementary surveys: (1) a sample of telephone-owning households reached by random digit dialing and (2) an area probability sample of non-telephone-owning (“nontelephone”) households reached by interviewers screening dwelling units for such households in a sample of tracts shown in the 1990 Census to have high proportions of nontelephone households.²⁰ Both samples include only households with at least one person under age 65. Considered together, the two samples are designed to constitute a probability sample of all age-eligible households in the United States.²¹ The Urban Institute contracted with Westat, a well-known survey research firm, to design the samples, conduct the interviews, and prepare the resulting data sets for analysis. All told, about 50,000 respondents within about 45,000 households were to be in the sample.

²⁰ To make the interview situation as comparable as possible in the two samples, NASF-I interviewers carried with them a cellular phone. When an eligible respondent gave permission to be interviewed, the interviewer used the phone to contact Westat’s telephone interviewing center, which then conducted the interview over the phone. Unfortunately, the sampling strategy for the nontelephone households turned out to be inefficient: Tracts that had high proportions of such households in 1990 no longer had that characteristic when screened in 1997.

²¹ People living in institutions, in shelters, on the streets, or on military bases were excluded.

2: Research on PRWORA: What Can Be Learned

Within each of the main samples, the following categories of households of special interest were oversampled:

- Households within each of the thirteen states designated for intensive administrative data collection (about 3,000 households in each state).
- Households with incomes 200 percent or less of the poverty line.
- Households with children age 18 or under.

Within households, respondents were selected according to the following rules:

- For households with children age 18 or under, the adult who was the primary caretaker of a randomly selected child under age 6 was chosen; in households with children between ages 6 and 18, the primary caretaker of a randomly selected child was chosen as respondent. When households contained children in both age ranges, separate adult caretakers were chosen when that was appropriate. The main intent of this rule was to obtain respondents who knew the selected children well enough to be able to answer questions about them in a knowledgeable way.
- In households containing no children age 18 or younger, an adult was randomly selected as respondent.
- In households with children and adults who were not parents of the children present, one of the nonparents was selected as a respondent.
- In households with large numbers of adults, more than one adult was selected.

In NSAF-I, the interviewing proceeded as follows: A short screening interview identified the eligibility of the household in terms of composition and income and identified which members would be subjects for a longer interview. All the longer interviews contained sections on employment, income, program participation, and health insurance and obtained information on the health status of the respondent as well as other adults in the household. Those who were selected as primary caretakers were asked additional questions about the selected children.

Interviewing for NSAF-I began in early 1997 and was completed in November 1997. The interviewing period took longer than initially planned because extra effort (and funds) had to be

2: Research on PRWORA: What Can Be Learned

expended in order to achieve a high response rate. The final weighted response rates²² (Urban Institute 1999f) were computed as shown in table 2–1.

Table 2–1 shows the weighted response rates separately for households with children and all other households. Rates are shown for the screening interview, for the extended interview conditional on completing the screening interview, and for the combined screening and extended interview. Because critical data are obtained only when screening interviews are followed by completed extended interviews, the substantively important response rates are 64.5 percent for households with selected children and 61.7 percent for households without selected children. Those response rates mean that slightly more than one-third of sampled households with children and almost two-fifths of sampled households without children were not included in the resulting data set.

Response rates varied by geographic locations of the respondents. For example, the highest response rate for households with selected children (74.5 percent) was obtained in Minnesota, whereas the lowest (55.5 percent) was obtained in New Jersey. A higher response rate (83.3 percent) was obtained in the area sample of nontelephone households, and a lower rate (64.6 percent) for households contacted in the telephone sample.²³

Although the response rates of NSAF-I are not high, they are typical of well-run national telephone surveys.²⁴ For example, a 1995 national telephone survey of adults ages 25 to 75, which screened households for eligibility and followed up with a mailed questionnaire, achieved an overall response rate of 61 percent; that is, 61 percent of eligible respondents completed both the screening telephone interview and returned a long mailed questionnaire.²⁵ In an article summarizing response

²² Because the survey is complex, with different elements of the population being chosen with various probabilities, the response rate was weighted to bring each segment sampled into line with those probabilities and with the sizes of the segments involved. (Sampling probabilities are varied to produce desired sample sizes of critical categories. Hence they are part of the design, being determined by the researchers' interests in subgroup sizes.) For example, the nonresponses in a segment that was oversampled would count less in the final weighted response than a segment that was undersampled. The weighting objective is to arrive at a rate that reflects what the NSAF-I fieldwork would have achieved had it been a simple random sample. The weighting scheme produces rates that are one to two percentage points higher than the response rates obtained from unweighted data.

²³ Response rates for telephone surveys are usually lower than for face-to-face surveys.

²⁴ Better response rates are achieved by sample surveys run by the Bureau of the Census, presumably because respondents believe that responding to government surveys is more important.

²⁵ Statistics are from the Survey of Midlife in the United States (MIDUS), conducted for the MacArthur Foundation Research Network on Successful Midlife (Brim and Featherman 1997). It should be noted that the MIDUS survey did not oversample poor households, nor did it include adults under age 25.

2: Research on PRWORA: What Can Be Learned

rates achieved in twenty-nine large national surveys conducted by nongovernment survey organizations in the 1990s, Massey et al. (1998) found that the average response rate was 62 percent, with about half of the response rates ranging between 60 and 70 percent.

Table 2–1. Weighted Response Rates for NSAF-I

Types of Respondents	Percentage
Response rates for households with children	
Completed screening interviews	77.8
Completed extended child interview after screening	84.1
Completed both screening and extended interview	65.4
Response rates for households without children	
Completed screening interviews	76.6
Completed extended adult interview after screening	79.9
Completed both screening and extended interview	61.7

Note: Data combine telephone and area samples.

Source: Brick and Flores-Cervantes (Washington, D.C.: Urban Institute, 1998).

Much better response rates, however, are achieved in face-to-face interviewing of low-socioeconomic status (SES) populations. For example, significantly higher response rates have been achieved in face-to-face surveys of low-income populations by MDRC survey contractors: In eleven surveys of welfare participants, eight had response rates of 80 percent or greater, and the lowest response rate was 70 percent.²⁶ As will be shown below, MDRC achieved an overall response rate of 79 percent in the first round of interviewing in the Urban Change project. The nontelephone portion of NSAF-I also had notably higher response rates: NSAF-I response rates for completed interviews with adult-only households were 80.5 percent for nontelephone households, compared with 61.0 percent for comparable households that had telephones.

Response rates matter in any survey. For surveys whose subject matter is closely related to the probability of responding, however, response rates can be critical: “Average” may not be good enough. For almost all surveys, response rates are lower for minorities, the poor, the poorly educated, and young adults, all of whom are population elements of special interest in NSAF. It is almost a certainty that as a group, nonresponding households in NSAF have those characteristics to a greater extent than those who responded.

The Urban Institute funded Robert Groves and the Survey Research Center at the University of Michigan to interview nonresponding households with a shortened version of the NSAF

²⁶Memorandum from Greg Hoerz of MDRC dated April 23, 1998. The surveys were done under subcontract to several different survey organizations.

2: Research on PRWORA: What Can Be Learned

interview (Urban Institute 1999e). Although that study had a low response rate,²⁷ the resulting data provide some critical information on the characteristics of the nonresponders. Apparently, compared with respondents, nonresponders to the screening interview tended to be of lower SES and nonrespondents to the extended interview tended to be of higher SES. Of course, these differences could be measured in the shortened version of the NSAF-I questionnaire, but important unmeasured differences may exist. This is of concern, because unmeasured differences might lead to quite different responses and hence bias. Groves' report (Urban Institute 1999e) also analyzed differences among respondents who participated on the first contact compared with those who responded after repeated contacts, finding systematic differences according to how much effort it took to achieve cooperation. Clearly some biases in response exist. Although the biases are not large, they do indicate that NSAF-I underrepresents critical demographic groups. For some purposes, these differences are ignorable, but for other purposes, they may be critical. In any case, users of the NSAF data should be aware of the potentials for bias and qualify their findings accordingly.

The technical research undertaken by the Urban Institute on the NSAF-I response problem is commendable as highly responsible. The NSAF-I data set will be analyzed by many social scientists, who need to be aware of the biases introduced by nonresponse, and the Urban Institute has taken pains to describe those biases.

No widely accepted standards have been adopted for judging whether the response rate for a survey is acceptable. Clearly, the higher the response rate, the better, but whether a particular response rate is acceptable is a judgment call. Affecting such judgments are such factors as the policy importance of the survey, how the typical response-rate patterns are related to the central topics of the survey, and whether the survey mode and design adhere to standards concerning best survey practices. In the case of the NSAF, the response-rate issue is clearly important. The central survey topics are politically salient, and the response rates are lowest for population subgroups that are of central interest. Arguably the telephone survey mode used may have been less expensive, but face-to-face interviews typically have higher response rates. At best, these considerations make NSAF response rates worrisome; at worst, they support a judgment that the survey may not be valid.

Consistent with best practices, NSAF will attempt to compensate for potential nonresponse bias by devising weighting schemes for the data set²⁸ that reflect the probabilities of nonresponse within population subgroups (Urban Institute 1999c.) The rationale underlying such weighting

²⁷ The survey received a 48 percent response rate. However, recall that this is a survey of households who repeatedly were contacted in NSAF-I and either did not initially respond or refused, sometimes repeatedly.

²⁸ Weights also will be constructed reflecting the different probabilities with which elements of the sample were selected. Those sampling weights are based on the sample design used. Westat also plans to construct "poststratification" weights, which are designed to bring the sample into line with known characteristics of the U.S. population as determined in the Census and updated by the CPS.

2: Research on PRWORA: What Can Be Learned

schemes is that people or households interviewed from among highly nonresponding groups can serve as proxies for those who did not respond from those groups. For a simplified example, if 25 percent of below-poverty-level inner-city households with children responded to the survey, the assumption is made that those who did not respond in that subgroup can be represented by those who did. Accordingly, if the average household income for responders in that population subgroup is \$6,000, that average is also attributed to the nonrespondents. Hence, responding households are given a weight of four in any calculations that include that population subgroup.

How well such weighting schemes compensate for nonresponse biases depends on how good the assumptions concerning proxies are.²⁹ Those assumptions, in turn, are affected by how much is known about how nonresponse varies by respondent features that can be identified in the survey. In most surveys, those assumptions cannot be tested. NSAF-I is somewhat better off than the usual survey because of the information contained in Groves' research (Urban Institute 1999d). However, even if the patterns of nonresponse within subgroups are well understood, it can never be known how well the responding households used as proxies represent the nonresponding households. Indeed, about the best that can be said about weighting schemes designed to compensate for nonresponse is that they are usually better than having no scheme at all, the latter being a strategy that is based on the clearly false assumption that nonresponders collectively are identical to the overall average responder.

In addition, the Urban Institute staff have contrasted NSAF-I's weighted findings³⁰ with those of the Current Population Survey (CPS) on comparable data items (Urban Institute 1999a). Comparisons of family composition, work experience, earnings, income, and poverty status by key demographic characteristics show a great deal of agreement between NSAF-I and the CPS. Some minor disagreements were found: Compared with CPS data, NSAF-I had larger proportions reporting full-time, full-year work; weekly earnings in the NSAF-I were slightly higher for women and slightly lower for men; and poverty rates were higher in NSAF-I (14.8 percent, compared with 14.1 percent in the CPS). Given that the instruments used were not exactly comparable, the discrepancies found are certainly minor. Although the comparability found is heartening, it is not to be regarded as definitive. The CPS is arguably the best comparison standard available for some outcome measures, but it also has defects. It has consistently underestimated welfare and Food Stamp Program participation levels when compared with administrative data, and the gap has been growing in recent years (Bavier 1998). For example, in 1997 the CPS estimates of welfare participation were only two-thirds of what was reported in administrative records. Comparisons of Survey of Program Dynamics data for welfare recipients against data available in administrative benchmarks would be helpful in assessing the quality of the data. In addition, the comparisons are limited: Important response biases on respondent characteristics may not be captured by the surveys.

²⁹ The weighting scheme used in NSAF-I was not assessed for this paper.

³⁰ Poststratification weights were applied to NSAF-I using home ownership and educational attainment variables in the CPS. Sensitivity analysis indicated that the poststratification weights are a minor factor in the CPS's comparisons on other demographic and socioeconomic variables.

2: Research on PRWORA: What Can Be Learned

The Urban Institute staff has thoroughly investigated the psychometric properties of scales measuring adult and child functioning (Urban Institute 1999d), showing them to have generally adequate to good internal reliability. The scales also correlate in “expected ways” with family structure and socioeconomic characteristics. That is, adults and children who are poor and in female-headed households tended to be worse off compared with those having opposite characteristics. Because many of the instruments were adapted from existing scales, attempts were made to compare NSAF-I findings with other studies. Unfortunately, differences in the populations sampled and in the scales used made such comparisons difficult to evaluate.

Urban Institute and Westat staffs have done considerably more than the usual sample survey staffs in conducting technical research to best counteract and compensate for potential nonresponse bias in the first NSAF survey. Nevertheless, whether those efforts are sufficient is likely to be an open issue. Consequently, analysts using the data set should be aware that those compensatory efforts may not have been successful and interpret findings accordingly. Although some strong confidence may be given to NSAF-I sociodemographic findings, more caution should surround the use of measures of child well-being, for which tests of potential response bias may not exist. Accordingly, it is important that the documentation accompanying the NSAF-I data set distributed for public use contain detailed discussions of how the data are to be used responsibly. The number of cases receiving welfare is relatively small in each state, however, making it unlikely that the effects of welfare reform can be discerned for all but major differences.

Recall that the main purpose of the NSAF-I survey was to provide good descriptions of how the populations of the thirteen selected states and the nation were faring before PRWORA went into effect. As far as the nation as a whole is concerned, some of the information that can be obtained from NSAF-I is already available from other national surveys, such as the CPS and the SIPP, both conducted by the Bureau of the Census. The unique contributions of NSAF-I are its large samples for each of the thirteen selected states, the existence of administrative case studies for each of those states, and its attention to important substantive topics, especially measures of the well-being of children, which are not collected in either SIPP³¹ or CPS.

Considered by itself, the NSAF-I study can support a variety of interesting analyses. It can be expected that much attention will be given to interstate comparisons. To an extent limited by the fact that only thirteen states are being studied, NSAF-I can be the vehicle for analyses of interstate differences in pre-PRWORA welfare systems and the conditions of the low-income populations in those states that are related to those systemic differences. However, it should be recalled that in 1996, AFDC was already in transition in many states: The AFDC waivers obtained in the 1990s had changed AFDC in radical ways in states such as Wisconsin, Michigan, Massachusetts, and Florida. The measures that are unique to NSAF-I likely will support interesting analyses, many centering on the issue of how children living in poverty households differ from those living in more affluent households.

³¹ Note, however, that the Survey of Program Dynamics will include questions on child well-being.

2: Research on PRWORA: What Can Be Learned

Of course, the NSAF-I data set is a cross-sectional survey. Causal inferences can be made, but they cannot be definitive or highly credible. To find that children in AFDC families in Mississippi are worse off than comparable children living in Minnesota can scarcely be attributed with great confidence to the less generous Mississippi welfare system because the two states differ in so many other ways.

Data collection for NSAF-II was completed in 1999. It was designed using the same sampling strategy and much the same interview schedule as the NSAF-I, with the addition of questions concerning experiences under PRWORA, use of the Earned Income Tax Credit, and changes to questions arising out of NSAF-I experiences. About two-thirds of the telephone numbers used in NSAF-I have been contacted again along with a new sample of telephone numbers. A similar strategy was followed for the nontelephone households. It is estimated that about two-thirds of the resulting sample was obtained from the NSAF-I telephone numbers and addresses. Although no special measures were taken to reach the same respondents as in NSAF-I, many of the respondents reached in recontacted households were the same people interviewed in NSAF-I. The NSAF-II sample design has some of the statistical advantages of a panel study, especially greater precision in detecting changes between NSAF-I and -II. Final NSAF-II response rates have not yet been calculated, although there are indications that they will be somewhat higher than those for NSAF-I.

Considered as a separate survey, NSAF-II certainly will be of interest in its own right. However, the major interest will be in the search for differences between the findings of NSAF-I and NSAF-II and relating the differences found to the workings of PRWORA. Despite the disclaimers from Urban Institute and Westat staffs that the two surveys together cannot be considered as properly supporting such causal inquiries, many analysts no doubt will try to draw such inferences. In 1999, NSAF-I public use data sets began to be released, and NSAF-II public-use data sets began to be released in 2000.

As is the case with any public-use data set, no way exists to prevent any given use of the data sets, no matter how flawed the analytic approach may be. The main corrective to misuses of the NSAF data set doubtless will be the intensive scrutiny and criticism that will be given to all analyses by the community of policy analysts and evaluation experts.

The essential feature of a before-and-after design is that it relies on differences between measures taken before and after a program is put into place as the basis for inferences about the effects of the program in question. The major problem with that design is that the differences detected may be a result of the program as well as anything else that transpired in the time between the two sets of measures. For measures that may be affected by changes occurring independently of the program, it is not possible to sort out definitively the amount of change caused by the program from the change resulting from other causes. It is possible that the employment levels and earnings of low-income households may show an increase between NSAF-I and NSAF-II, but that increase

2: Research on PRWORA: What Can Be Learned

may stem from, say, changes in the labor market. Even if employment levels and earnings decline between NSAF-I and -II, that may not mean that PRWORA was ineffective, because that program may have prevented a more precipitous decline. The obverse also may be true: The increase in earnings and employment might have been greater without PRWORA.

The limitations of NSAF's before-and-after design mean that plausible causal inference from the resulting data sets must draw heavily upon information gathered from other sources. For example, any inferences from NSAF about the effects of PRWORA on employment must take into consideration trends in employment generally in the thirteen states, especially for low-wage jobs. Inferences about the effects of time limits on family well-being need to be bolstered by information on how the time limits were administered in each of the states. (Of course, other factors that might affect those topics need to be considered as well, such as local labor market conditions.) Although all such analyses will be fragile, the best (and most credible) will be those that test and rule out alternative explanations for the outcomes under discussion. These considerations enhance the importance of the thirteen administrative case studies and the state database, two other potential sources of information about changes accompanying the implementation of PRWORA.

Bureau of the Census: Survey of Program Dynamics. For about fifteen years, the Bureau of the Census has been conducting a national longitudinal household survey known as the Survey of Income and Program Participation (SIPP).³² The survey concentrates on issues involving household income, employment, and participation in a wide variety of welfare programs as well as associated topics, such as physical and mental health and household use of child care services. SIPP is based on an area probability sample, with oversampling of households living in low-income areas. About 37,000 households are interviewed in every four-month period. The sample consists of four equivalent panels, each of which is interviewed repeatedly over a period of four years. Interviews every four months query each household about employment, income, and program participation for each month of the preceding quarter and collect data on changes in household composition. Special supplemental interviews are administered from time to time; they cover topics such as child care arrangements or physical disabilities. SIPP is the major source of information about household patterns over time in the relationships among employment, poverty, and welfare receipt and the events that change them.

The PRWORA legislation contained a provision that directed and funded the Bureau of the Census to extend data collection from the SIPP households recruited in 1992 and 1993 through 2001, appropriating \$70 million for that purpose.³³ This extension of SIPP, now called the Survey of Program Dynamics (SPD), will provide longitudinal data on about 19,000 of the approximately

³² This section is based on Weinberg et al. (1997) and questionnaires to be used in SPD furnished by the Bureau of the Census.

³³ The initiative for this provision stemmed from long-range planning activities in anticipation of welfare reform funded by HHS and the U.S. Department of Agriculture, which supported the Bureau of the Census in preparing plans for extending SIPP some years before the enactment of PRWORA.

2: Research on PRWORA: What Can Be Learned

30,000 panel households for four to five years before and for five years after the enactment of PRWORA.

The original plans called for SPD to start in 1996. Because PRWORA was not passed until August 1996, the survey could not be started until 1997, causing a significant gap in data collection for the 1992 SIPP panel, which had last been interviewed in January 1995. To fill the gap, a “Bridge Survey” was designed to gather data on income and program participation in the missing years. The Bridge Survey was essentially a modified version of the 1997 March supplement of the CPS and collected information for the missing period. As a result, data for 1995 and 1996 for the 1992 panel and for 1996 for the 1993 panel are not as detailed as in previous years for SPD households and are based on a longer period of recall. Specially prepared questionnaires also will be used with subgroups of particular interest, such as adolescents and married adults.

The Bridge Survey was followed in 1998 by interviews using questionnaires tailored for SPD purposes. SPD interviews will proceed on an annual basis until 2002. Each annual interview will cover the previous year (for example, information on 2000 will be collected in the 2001 interview). Note that SPD data for the post-PRWORA years will not be as detailed as for the pre-PRWORA years. New topics, however, especially measures of child well-being,³⁴ were added beginning with the 1999 SPD.

Unfortunately, Congress did not appropriate enough funds to interview all the households in the 1992 and 1993 panels. The Census Bureau thus reduced the 1998 SPD sample size to about 18,500, retaining all the households residing in low-income areas and sampling a smaller share of those living in more affluent areas.³⁵

Despite the budget and data problems, SPD will provide data for a national panel study of households repeatedly interviewed over about a decade, the years more or less evenly divided into pre- and post-PRWORA years. The long pre-PRWORA period is intended to serve as a firm baseline measurement against which to contrast the changes that might be wrought by PRWORA. The long post-PRWORA interviewing period will track households through the early implementation stage of PRWORA and into the first few years of more settled forms of PRWORA.

³⁴The measures closely resemble those being collected in the NSAF. Child Trends, Inc., is responsible for item development for both the Bureau of the Census and the Urban Institute. Because the items were added only beginning with the 1999 SPD wave, no pre- and post-PRWORA comparisons will be possible.

³⁵ The following subsamples will be included: (1) all households at 150 percent or less of the poverty line, (2) all households with children between 150 percent and 200 percent of the poverty line, (3) 90 percent of households with children at 200 percent of the poverty line or above, (4) 82 percent of childless households between 150 percent and 200 percent of the poverty line, and (5) 27 percent of childless households at 200 percent of the poverty line or above.

2: Research on PRWORA: What Can Be Learned

The Bureau of the Census does not intend to undertake any sophisticated analyses of the effects of PRWORA. The SPD data set will be released for public use, and effectiveness analyses will be undertaken by social scientists in research universities and in the major research institutes.³⁶

Maintaining the cooperation of households throughout an extended period of repeated interviewing is a difficult task. In both the 1992 and 1993 panels, about one-fourth of the households who were initial participants were no longer cooperating by the end of their initial four years of participation. An additional 18 percent attrition rate was experienced in the Bridge Surveys described above. Attrition rose again in the 1998 SPD survey, to 50 percent. The response rate did not go below 50 percent in 1999, likely because the Census Bureau made special efforts, including incentive payments, to bring some nonrespondents back into the sample. Given unchanging survey strategies, additional attrition is likely in future SPD waves. The final SPD data set, incorporating all the data collected, can be safely expected to include less than 50 percent of the households who gave interviews in the 1992 and 1993 panels. The result is that analyses requiring uninterrupted measures across the full decade of SPD interviews will be based on data sets that have experienced perhaps as much as 75 percent attrition.

Although cutting back on the sample size for SPD does not introduce any potential bias, the fairly sizeable attrition likely does. Because something is known about each of the households who stopped cooperating, the nature of the bias has been investigated. It appears that nonresponse was greatest among the poorest respondents, precisely the group about whose condition under PRWORA generates the most interest. We can gain some insight into the nature of the biases that the correlates of attrition introduce in the later interviews by examining the responses of households of similar characteristics who cooperated. Although weights can be calculated based on the data on nonresponding households in an attempt to offset attrition biases, the success of the weighting scheme in dealing with selection biases will be unknown.

Concerned about the high nonresponse rates, the Bureau of the Census conducted the Exploratory Attrition Study in early 1999 to ascertain the feasibility and cost of obtaining cooperation from nonrespondents. A sample of 406 households, composed partly of nonrespondents to the original 1992 and 1993 SIPP along with those who did not respond to the Bridge Survey, were traced to their current addresses; when contacted, interviews were attempted. Three levels of monetary incentives were used: \$0, \$50, and \$100. Although the total response rate was 37 percent, those offered \$100 as an incentive had a 44 percent response rate, compared with 29 percent for those offered no payment.

Encouraged by these results, the Bureau of the Census has asked Congress to appropriate funds to attempt to reach and convert to cooperation all the nonrespondents who were at or below

³⁶ Such efforts are already underway. Financed by grants from several private foundations, Thomas MaCurdy of Stanford University's Hoover Institute is planning interrelated analyses of SIPP, SPD, NSAF, and the Panel Study of Income Dynamics that will focus on the impact assessments of PRWORA and use sophisticated econometric modeling.

2: Research on PRWORA: What Can Be Learned

200 percent of the poverty line when they had been cooperating. Using the \$100 incentive payment and a \$40 incentive for each subsequent completed interview, the Bureau of the Census expects to achieve a response rate of close to 63 percent by the 2003 SPD interview. The current budget contains about \$1 million to start on the project, with prospects for an additional \$18 million to complete the process. Of course, it is not known whether the projected 63 percent response rate can be achieved. Certainly, there would be some improvement, perhaps even greater than expected. The value of the SPD data set would certainly be enhanced.

The Bureau of the Census also is investigating the feasibility of linking the 1992 and 1993 panels to Social Security Administration records of employment and earnings. Because social security numbers are available both for those who consistently responded and for those who did not, it will be possible to have that data for all who initially enrolled and for the entire period pre-and post-PRWORA. Assuming that the linking procedures can be reasonably successful, employment and earnings data would be available for 80 percent or more of the SPD panels. The effort would be a significant enrichment of SPD.

The information obtained from households beyond 1995 for the 1992 panel and 1996 for the 1993 panel is not of the same character as that obtained prior to those dates. In SIPP, households are interviewed every four months; in SPD, data are obtained annually. Information recalled over a four-month period is almost certainly more accurate than similar information obtained by recall over a full year. In short, the post-PRWORA information is likely to be more subject to recall errors than the pre-PRWORA information is.

Attrition and data differences thus lower the value of SPD data sets for analyzing the impact of PRWORA. Without corrective action to raise response rates, the percentage of cases with full sets of interviews may well be around 25 percent, too low by any standards. Note also that the planned efforts to improve the response rate will not improve the numbers for full panel information. The newly cooperating respondents all will have gaps consisting of the interviews missed while not cooperating.

Response rate problems are even more serious for some specialized surveys conducted periodically as part of the SPD. For example, the 1998 Adolescent Self-Administered Questionnaire achieved a 60 percent response rate, but because the rate is calculated based on an overall SPD response rate of 50 percent, the findings are based on adolescents from just 30 percent of the original households.

Putting possible data quality problems aside, what can be learned from SPD about the impact of PRWORA? Even under the conditions that SPD data are excellent, the study is still a before-and-after design, differing from NSAF primarily by having more information on the life trajectories of households for both periods. That information advantage is considerable, but it does not counteract the basic deficiency of such designs, namely an inability to distinguish between the effects of PRWORA and the effects of other trends occurring at the same time.

2: Research on PRWORA: What Can Be Learned

MDRC: The Project on Devolution and Urban Change. The Manpower Demonstration Research Corporation has conducted many of the better-known state welfare waiver experiments over the past decade or so, accumulating an impressive record of research accomplishments.³⁷ MDRC's Project on Devolution and Urban Change (UC) combines many of the features of SPD and NSAF, although it focuses on cities rather than states or the nation as a whole. It also differs from the other studies in its greater reliance on administrative data for measuring employment and earnings and in having an ethnographic component. Of all the research reported in this paper, UC is more sharply focused on TANF, being concerned more with recipients of welfare than with poor households generally.

The Urban Change project currently has a projected budget of about \$13 million and is funded through private foundation grants. It is anticipated that some of the project components will turn out to be more expensive than projected: The final budget could be closer to \$16 million. In keeping with MDRC policy, the resulting data sets eventually will be released for public use.

UC is planned as a set of community studies involving four cities: Cleveland, Philadelphia, Miami, and Los Angeles. The choice of those cities was heavily influenced by the openness of the political jurisdictions to MDRC access to administrative data. In each of the four cities, several research operations are planned (some of which are underway):

- Ethnographic studies in three poor neighborhoods in each site, in which about a dozen single mothers will be interviewed about their experiences with the welfare system and their work experiences.
- Implementation studies of TANF in each site, which will be based on documents, interviews with agency officials, focus groups with welfare workers, messages given to clients by agencies, and media stories about local TANF issues and activities.
- Institutional studies of the impact of TANF on for-profit and nonprofit institutions, such as schools, police departments, banks, local businesses, and private social service agencies.
- Neighborhood indicator studies using existing social indicators (for example, fertility, mortality, and crime). The indicators will be used to identify and track changes in poor neighborhoods over time.
- Sample surveys of single mothers, which consist of a 1998 survey of 1,000 mothers who were on AFDC in May 1995 and who were living in high-welfare or high-poverty neighborhoods, and a second survey (of the same size and composition) of mothers enrolled

³⁷ This section is based on MDRC project design papers (Michalopoulos et al. 1997; Quint 1997; Coulton and Verma 1997; Edin 1997; and various communications with Gordon Berlin and Charles Michalopoulos.

2: Research on PRWORA: What Can Be Learned

in TANF at a yet-to-be-determined post-PRWORA point. The interviews covered employment, program participation, income, fertility, and measures of family and child well-being.

- Longitudinal administrative data covering successive cohorts of nonelderly AFDC/TANF and Food Stamp Program enrollees from 1992 through 2002. Earnings records from employers' unemployment insurance reports will measure employment and earnings in covered employment. AFDC/TANF and Food Stamp Program records will be used to measure participation in those programs.

The six research projects that make up UC are designed to complement each other to produce a rich description of the systemic changes in welfare sparked by PRWORA and accompanying changes in the circumstances of poor households. Several are unique to UC. For example, none of the other major welfare reform studies contain an ethnographic component, and UC is the only study that will provide information on how the members of the "street bureaucracy" of welfare agencies interpret their roles and behave toward welfare clients.

Especially characteristic of MDRC is the important role given to administrative data. That data will be used in a multiple-cohort time-series design for estimating effects of PRWORA. That design is critical to UC estimates of the effects of PRWORA. Accordingly, the focus of this review is on how UC plans to use administrative data.

Although the details are still being worked out, the plan for using the multiple-cohort time-series design is as follows:

AFDC/TANF and Food Stamp Program administrative data will be used to define cohorts of people enrolled in those programs. The UC project will restrict its concerns to the nonelderly people on the Food Stamp rolls. A cohort is a group of people or households defined as being in some identifiable state at some point in time; hence, a 1992 AFDC/Food Stamp cohort can be defined as everyone on AFDC and/or Food Stamp rolls in that year.³⁸ They remain in that cohort until the end of the research period, and data for that period constitute the cohort data file. The research project runs from 1992 to 2002.

Using the employer files of state employment security agencies, each cohort's employment and earnings will be obtained for each quarter in each year from the start of the cohort to the end of the research period. Administrative records from AFDC/TANF will be used to track enrollment in AFDC/TANF and receipt of income maintenance. The earliest cohorts will provide data on pre-PRWORA earnings, employment, and welfare reciprocity as well as post-PRWORA measures. Cohorts defined after 1997, of course, will only reflect experience under TANF. AFDC/TANF

³⁸ The research team also is considering defining each cohort as those who enrolled in AFDC/TANF or the Food Stamp Program over a given period of time. In addition, no decision has been made on how long the initial period should be (for example, a year, a half year, or some shorter period).

2: Research on PRWORA: What Can Be Learned

administrative data also contain information on some of the demographic characteristics of members of enrollees, including household composition, race, and marital status. The separate administrative data sets can be linked through such common unique identifiers as social security numbers, birth dates, and names.

The multiple-cohort design has several desirable features. First, the resulting data will provide much detail on both pre- and post-TANF participation and labor force behavior and thus a firm basis for characterizing those periods. Second, the data are not based on recall but are recorded close to the time of occurrence; for example, the data on earnings are based on payroll records and reports that employers turn in every quarter with their employment tax returns. Third, the data can be used to show how PRWORA may have changed the composition of TANF clients compared with the AFDC population. For instance, it may turn out that people with earnings potential may prefer to seek employment rather than enroll in TANF, lowering the average educational attainment of new enrollees in TANF compared with new enrollees for AFDC. Conversely, the more generous earnings disregard under TANF might encourage the working poor to enroll.

The multiple-cohort time-series design also has problems. The data set is confined to households enrolled in AFDC/TANF, Food Stamp Program, or both. Although this set includes a large percentage—perhaps the majority—of the poor, many eligible households are not enrolled in either program. The more inclusive set consists of those enrolled in the Food Stamp Program, but the size and composition of this group can be affected by changes in food stamp eligibility criteria, which were changed in PRWORA and may change again in the future.

The earnings and employment data set reflects only employment and earnings that are covered by the unemployment system. As several researchers (notably, Edin and Lein 1997) have shown, many AFDC recipients participate in the “underground economy” in employment that is not recorded in those records as well as receive income in the form of gifts from others.³⁹ It is problematic whether the more generous income disregards put in place under most state TANF plans will reduce participation in the underground economy or increase the reporting of such income. A related problem is that the administrative data sets will not cover people for any portion of the study period in which they lived out of state. Of course, in-migrants and people leaving the state will constitute relatively small groups. MDRC hopes to reach by telephone those who have left the study communities.

Although administrative data have only minimal response problems, they are not entirely without quality problems. Linking several data sets rarely can be done without error. Some data

³⁹ The two surveys and the ethnographic studies may provide estimates of the extent of such income sources, but it is unlikely that those estimates will be able to supplement the administrative data except for those who are included in the survey samples. Underreporting of “under-the-table” earnings and gifts is a problem in most surveys. Edin and Lein (1997) report that they were able to obtain such information only after establishing strong rapport with respondents.

2: Research on PRWORA: What Can Be Learned

elements may be missing. That said, administrative data typically suffer from considerably fewer data quality problems than survey data.

The multiple-cohort time-series design has only limited ability to separate TANF effects from those of other historical events or trends. For example, the several cohorts will come up against time limits at different time points. If the effects of time limits are approximately equal across the different cohorts, one can be confident that such effects are independent of historical trends and events. However, trends whose influences cannot be identified may occur post-TANF. For example, the effects of a trend change (for better or worse) in the employment prospects for people seeking low-skill jobs likely cannot be clearly distinguished from the effects of TANF.

The two sample surveys of welfare participants are another important project component. The surveys are planned as before-and-after studies and are intended to measure outcomes that are not available in administrative data, such as measures of family functioning, health status, child well-being, and so on. The first survey, completed in 1998, consisted of samples in each of the four cities of 1,000 single mothers who were enrolled in AFDC in May 1995 and who lived in selected neighborhoods high in poverty and welfare participation. Much of the interview was focused on respondents' experiences over the year prior to the interview. The overall response rate to the 1998 survey was 79 percent, clearly a high value. The second survey will be of 1,000 single mothers who were enrolled in TANF after PRWORA and who live in the same neighborhoods. The second survey focuses on experiences under TANF as well as contemporary issues. Both surveys were designed as face-to-face interviews.

An important feature of the UC surveys is that their samples are drawn from the AFDC/TANF administrative data files and therefore can be linked to those files and employment and earnings files. That feature considerably enriches the resulting data set by adding employment, earnings, and welfare participation data from the administrative files for periods before and after the time of interview.

Comparisons across the two UC surveys may present problems. Each of the UC surveys is drawn from different populations. It is quite possible that the May 1995 AFDC clients differ in important ways from the post-PRWORA TANF clients; TANF applicants self-select themselves, may face different eligibility rules, and hence may be quite different from AFDC clients. Comparisons across the two surveys will require statistical adjustments.

HHS ongoing studies: The Child Impact Waiver experiments. Some of the Department of Health and Human Service's waiver experiments put in place during the 1980s had been completed by the time PRWORA was enacted, but many of those started in the 1990s were still

2: Research on PRWORA: What Can Be Learned

underway.⁴⁰ The states were given the option of continuing the latter, and more than a dozen states elected to continue their experiments.

Although policy concerns about the impact of TANF centered on whether welfare clients would be weaned from dependency on benefits and moved into the labor force, strong interest also was generated about TANF's effects on children. As single mothers became employed, young children would move into out-of-the-home child care with unknown effects on their well-being. Furthermore, the new time limits and sanctions for noncompliance raised questions about what would happen to children when benefits were reduced or ended. Unfortunately, not much could be learned about effects on children from the waiver experiments. Almost all the waiver experiments concentrated on outcomes such as adult employment and earnings. Data collection instruments typically contained few variables concerned with outcomes for children.

HHS decided to make some of the best continuing waiver experiments more relevant to TANF concerns by augmenting them with special data collection efforts adding child outcome variables. To make the augmented data collection efforts comparable across experiments, HHS issued invitations for state participation in a planning phase, the product of which would be a standard data collection instrument covering outcomes for children. About twelve states agreed to participate. After the planning phase was over, the participating states were asked to submit applications for support for augmenting their waiver experiments. Out of the eleven applications submitted, five were awarded support: Connecticut, Florida, Indiana, Iowa, and Minnesota. The five states agreed to use survey instruments, mostly comparable across states, designed to capture the differences in child outcomes between experimental-group and control-group families. The Child Impact studies will concentrate on families with children between ages 5 and 12, although some studies also will study younger children or adolescents. (Because the surveys take place three to four years after random assignment, the studies will capture the effects of the welfare reform experience on children who entered the program between the ages of 1 and 9.) About \$12 million in federal funding has been allocated to the five waiver experiments through 1999, augmented by some state and private foundation funding.

Although the five waiver experiments are concerned with employment and earnings effects, a new emphasis, child outcomes, has been added. The child outcome measures common to all five waiver experiments are shown in the last two columns of table 2-2; the first two columns show, respectively, the expected direct effects of the waivers and the intermediate effects that are expected to lead to the anticipated child outcomes. Of course, child outcomes cannot be measured

⁴⁰ This section is based on proposals submitted by the states to HHS for funding to augment the ongoing waiver experiments (Connecticut Department of Social Services 1997; Florida Department of Children and Families 1997; Indiana Division of Family and Children 1997; Iowa Department of Human Services 1997; Minnesota Department of Human Services 1997). Additional information was obtained from comments on an earlier draft of this paper from Howard Rolston of the Administration for Children and Families, HHS.

2: Research on PRWORA: What Can Be Learned

retrospectively. Each of the five states will measure those outcomes⁴¹ for several years beyond the start of the experiments. The Child Impact surveys initially were planned to take place between 1997 and 2000. Table 2–3 shows programmatic features of the Child Impact Waiver experiments.

⁴¹ The absence of baseline measures of child outcomes means that the estimates of effects will not be as precise as for employment and earnings, for which baseline measurements can be used as covariates.

Table 2–2. Common Construct Measures for HHS Child Impact Studies

Target of Welfare Policies	Other Variables Likely to Be Affected by State Policies	Aspect of Child’s Environment Likely to Be Affected by Previous Columns	Child Outcomes
<p><i>Income</i></p> <ul style="list-style-type: none"> • Total income • Sources of income (mother’s earnings, father’s earnings, child support, AFDC, Food Stamp Program, SSI, foster care/adoption) • Stability of Income • Financial strain/material hardship 	<p><i>Psychological Well-Being</i></p> <ul style="list-style-type: none"> • Depression <p><i>Stability and Turbulence</i></p> <ul style="list-style-type: none"> • Foster care • Stability in child care • Stability in income • Number of moves of residence • Change in marital status or cohabitation • Why child not living with family 	<p><i>Child Care</i></p> <ul style="list-style-type: none"> • Type • Extent • Quality (group size, ratio, licensing, parent perception) • Stability • Child care calendar for past several years 	<p><i>Education</i></p> <ul style="list-style-type: none"> • Engagement in school (ages 6–12) • School attendance (all children) • School Performance (all children) • Suspended/expelled (all children) • Grades (ages 6–12)
<p><i>Employment</i></p> <ul style="list-style-type: none"> • Any vs. none • Health benefits through employment • Wages (hourly) • Hours of employment • Stability of employment • Education/licenses • Job skills (hard) • Multiple jobs concurrently • Barriers to employment (harassment, violence) 	<p><i>Absent Parent Involvement</i></p> <ul style="list-style-type: none"> • Whether child support provided • Paternity establishment • Frequency of contact with child <p><i>Use of Health and Human Services</i></p> <ul style="list-style-type: none"> • Food stamps • Medicaid (awareness, use, eligibility) • Child care subsidy (awareness, use, eligibility) • Access to medical care 	<p><i>Home Environment and Parenting Practices</i></p> <ul style="list-style-type: none"> • Child abuse/neglect (administrative data) • Domestic violence/abusive relationships • Family routines • Aggravation/stress in parenting • HOME (Emotional Support and Cognitive Stimulation Scales) 	<p><i>Health and Safety</i></p> <ul style="list-style-type: none"> • Hunger/nutrition (ages 5–12) • Rating of child’s health (ages 5–12) • Regular source of care (ages 5–12) • Teen childbearing (ages 14–17) • (all children) • Accidents and injuries (all children)
<p><i>Family Formation</i></p> <ul style="list-style-type: none"> • Nonmarital birth/marital birth • Child/family living arrangements • Marital status, whether married to biological or nonbiological father 	<p><i>Consumption</i></p> <ul style="list-style-type: none"> • % of income spent on child care and rent 		<p><i>Social and Emotional Adjustment</i></p> <ul style="list-style-type: none"> • Behavior problems Index (ages 5–12) • Arrests (all children) • Positive Behaviors/Social Competence Scale (ages 5–12)

Source: Author.

Table 2–3. Features of Child Impact Waiver Experiments

	Connecticut	Florida	Indiana	Iowa	Minnesota
Program name	Jobs First	Family Transition Program	Indiana Welfare Reform	Family Investment Plan (FIP)	Minnesota Family Investment Plan
Contractor	Manpower Demonstration Research Corporation	Manpower Demonstration Research Corporation	Abt Associates	Mathematica Policy Research	Manpower Demonstration Research Corporation
Year started	1996	1994	1995	1993	1994
Sample	Manchester and New Haven: 1,200 families in both experimental and control groups	Escambia County: 600 families in both experimental and control groups	Clustered state sample: 1,300 families in experimental group and 1,300 in control group	Nine counties: 2,150 families in experimental group and 1,050 in control group	Eight counties: 1,977 families in experimental group and 1,323 in control group.
Focal child(ren)	Randomly selected child age 5–12	All children ages 5–12	All children ages 5–12	All children ages 5–12	Randomly selected child age 5–12
Time limits	21 months	24 months out of each 60 month period	24 months for adults	None	None
Earnings disregard	100% disregard until family reaches poverty level	First \$200 and 50% of any additional earnings	No change from AFDC	100% for first four months and generous work allowances and disregard thereafter	Generous income disregards with payments continued until family reaches 140% of poverty level
Family cap	Yes: reduced additional payments	No	Yes	No	No
Child care eligibility	Extended	24 months after leaving welfare	Extended	Transitional child care extended to 24 months	No enhancement over AFDC
Sanctions	Graduated sanctions for noncompliance increasing with each violation	For noncompliance	For noncompliance	For noncompliance	For noncompliance
Child Impact survey date	2000	1998–1999	2000	1997–1998	1997–1998
Special features	Emphasis on maintaining fidelity	Emphasis on maintaining fidelity	PRWORA provisions applied to experimental group in May 1997	Controls switched to FIP in March 1997	Two experimental treatments. Emphasis on maintaining fidelity

Source: Author

2: Research on PRWORA: What Can Be Learned

Although the five augmented experiments concentrate on the same child outcomes, they differ considerably in their waiver provisions, as shown in 3–4. Three of the five have time limits. Four of the five have more generous income disregards (compared with AFDC), but even those four vary considerably, from the most generous 100 percent earnings disregard (up to the poverty level) in Connecticut to the less generous Florida provisions. One of the experiments, Indiana’s, is statewide, whereas all the others are conducted in one or more subareas within the states in question. Two states have family cap provisions, and three do not. All the states apply sanctions for noncompliance with waiver provisions, but only Iowa and Connecticut cut off benefits after repeated violations.

The relevance of the ongoing experiments to TANF depends heavily on the extent to which the experimental conditions match those adopted by the states as their TANF provisions. By and large, the matches are quite close: TANF provisions apparently followed quite closely the treatments used in the experiments.

In all five states, changes were made in the experimental conditions to bring them more closely into line with TANF. (The appendix to this paper describes the changes made.) Many of the changes appear to be quite minor in that they affect small numbers of participants or are small changes. For example, the only change made in Minnesota’s experiment was the imposition of a five-year time limit on the experimental group, a change that makes the experimental group subject to all the provisions of Minnesota’s TANF. The greatest number of changes was made in Indiana’s experimental group, but most of the changes appear to be quite minor.

In Iowa the changes are major ones: The experiment essentially was discontinued in March 1997, and both control and experimental families were subjected thereafter to TANF provisions. However, because TANF provisions are almost identical to the experimental conditions, the 3.5-year pre-TANF experiment is relevant to that state’s TANF conditions. Of course, measurements taken after March 1997, including child outcomes, cannot be analyzed as stemming from a randomized experiment.

Whether each of the five experiments tests conditions that are close enough to the relevant state TANF provisions is clearly a judgment call. My view is that the experiments in Minnesota, Florida, Indiana, and Connecticut are close enough⁴² but that the data collected in Iowa after March 1997 will not be useful.

The five waiver experiments began and continue in the context of turbulent times for public welfare. The national as well as the state and local mass media have given much space and time to public announcements and discussions of PRWORA and TANF. For state and local

⁴² Three additional waiver experiments—in Vermont, Texas, and Arizona—are being continued by HHS, although they are not participating in the Child Impact Waiver experiment. The waiver conditions in the three states are quite close to the state’s TANF provisions.

2: Research on PRWORA: What Can Be Learned

welfare agencies, the times have meant changes both in professional duties and in agency regulations and procedures. Under these conditions, maintaining the fidelity of experimental conditions is a challenge to the success of the experiments. Ideally, welfare personnel should treat families in the experimental group according to the conditions assigned to them, and participating families also should know the critical features of those conditions. These strictures apply not only to the families who are active cases but also to those who are not. Families who have dropped off the rolls should know that if they were to apply again, the experimental conditions would still govern. Similarly, those in the control group who have dropped off the rolls should know that AFDC provisions would apply should they return to the rolls.

In three of the experiments—Connecticut, Florida, Indiana, and Minnesota—the operational plans include special efforts to maintain experimental conditions, including periodic communications to participating families, separate welfare staff assigned to participating families, and subroutines added to MIS systems to prevent inappropriate actions. Of course, how successful those efforts will be is unknown.

What Will Be Learned?

This section describes what can be learned from the studies reviewed. It first considers the three prospective research projects, then discusses the Child Impact Waiver experiments.

The prospective research projects. The three prospective PRWORA research projects have already spent more than \$150 million, and millions more will be spent over the next five years. Assuming successful completion of data collection for the three projects, what will be learned from them? When released, the descriptive findings undoubtedly will be met with intense public and partisan interest in how the poor are faring under PRWORA. These studies will provide that desired information in great detail.

The studies will be able to provide descriptive findings on the following levels of aggregation:

- **National level.** The Urban Institute's Assessing the New Federalism project and the Census Bureau's SPD will provide findings concerning how American households were doing socioeconomically before and after PRWORA. For the nation as a whole, we will know whether earnings and employment of low-income households have improved or declined. We also will know whether children in low-income families are better or worse off in certain respects. The Urban Institute's state database also will provide detailed data on how the fifty states used the opportunities offered by devolution to design their particular versions of PRWORA. We will know which states offered generous income disregards, which adopted stricter time limits than those called for in TANF, and which put caps on payments when mothers on TANF had additional children.

2: Research on PRWORA: What Can Be Learned

- **State level.** Assessing the New Federalism will provide detailed information on each of thirteen states. The administrative case studies promise to provide rich information on how the states ran AFDC and what changes were made under TANF. Implementation issues also will be studied. NSAF's sampling design is intended to provide large enough state samples (about 3,000 households in each of the thirteen states) to describe how the circumstances of the poor changed after PRWORA. Comparisons of AFDC and TANF will be able to be made among the states and within each state. The Urban Institute will provide detailed descriptions of PRWORA and TANF plans for each of the thirteen states. The Census Bureau's SPD is not based on a large enough sample to provide data on more than the largest states—perhaps only New York and California—and even then, the samples are quite small.
- **Local level.** MDRC's Urban Change project is concerned with four cities and promises to provide detailed longitudinal descriptions of how several cohorts of poor households changed in employment and program participation over the historical periods pre- and post-TANF. UC will provide descriptions of how TANF was implemented in the cities as well as changes in several poor neighborhoods in all the cities. UC's findings will be especially valuable for learning about how the changes making up PRWORA are being interpreted in practice, with the obvious limitation that such findings are limited to the four urban places under study.

Taken together, the descriptive information resulting from the three research projects surely will be valuable. It will be of great interest to find out how various subgroups among low-income households are faring under PRWORA. For example, when TANF participants are cut off from support, either through sanctions or from coming up against time limits, what happens to them? Will job-seeking success be more difficult to achieve among teenage single mothers than among older mothers? Will TANF participants in urban neighborhoods characterized by extreme poverty have different experiences compared with participants living in small towns? Whether these analyses can be carried out, however, depends on whether the data sets contain enough households in the situations in question. For example, the NSAF surveys do not contain enough TANF participants in many of the thirteen states to sustain subgroup analyses in those states, although the UC study will likely support such analyses in each of the four cities being studied.

In general terms, the three prospective research projects will produce information bearing on the following topics:

- Employment pre- and post-PRWORA (whether adults are employed and the time patterns of employment, such as steadiness of employment).

2: Research on PRWORA: What Can Be Learned

- Welfare benefits pre- and post-PRWORA (participation patterns and amounts of cash and in-kind benefits received, including food stamps, Supplemental Security Income, housing subsidies, Earned Income Tax Credit, and other programs directed at the poor).
- Earnings from employment pre- and post-PRWORA.
- Sanctions pre- and post-PRWORA (benefit reductions through noncompliance and benefit loss resulting from time limits).
- Household composition pre- and post-PRWORA (marriage formations and dissolutions and births to welfare clients).
- Child care pre- and post-PRWORA (child care arrangements, some quality measures, turnover in arrangements).
- Health and health care pre- and post-PRWORA (self-ratings of health status, depression, medical care accessibility).
- Child well-being measures pre- and post PRWORA (school attendance, parenting quality, child behavior problems, food insecurity).

Conspicuously missing from the list above are findings concerning the impact of PRWORA and TANF. They are omitted because the before-and-after designs cannot support credible impact analyses. Nevertheless, it can be safely predicted that attempts will be made to estimate the effects of PRWORA and TANF because of widespread interest in policy circles and in the general public.

An obvious approach to impact analysis would be to exploit the longitudinal features of the data sets, contrasting individuals, cohorts, or political jurisdictions before and after PRWORA and TANF were implemented. Because there is reason to believe that many changes affected households during the research period in addition to welfare policy, any analysis must try to take such changes into account. Especially important are macro-level changes in the economy and changes in other programs, such as the Earned Income Tax Credit program. The changes occurring at the micro level may be more difficult to model and take into account. For example, increased earnings of partners and in-kind transfers may make it possible for TANF recipients to leave the program and join other households. Whether such nonwelfare changes can be adequately taken into account in the before-and-after analyses will be problematic. Indeed, it is likely that analysts may come to quite different impact assessments depending on how their statistical models are constructed.

2: Research on PRWORA: What Can Be Learned

The data sets produced by the three projects also can be analyzed as cross-sections—snapshots taken at particular points in time. A likely strategy will be to contrast households who face different TANF plans or to compare households who are TANF participants with comparable households who are not. In such analyses, the major obstacle to credibility is whether analysts have properly managed to model and adjust for selection biases.

A related approach might be to contrast the clients of different welfare plans on the state, county, or city level. The point of such analyses would be to discern how the features of different plans affect outcomes for households subject to them. The cross-sectional approach is especially fraught with potential for error. As discussed earlier, the TANF plans are bundles that are impossible to “unbundle.” It is unlikely that any pairs of jurisdictions exist whose TANF plans differ in just one or even two features. It will be difficult to make a strong case for any impact estimate based on such comparisons.

Although statistical modeling cannot make up fully for the inherent weaknesses of before-and-after designs, it appears to be the best strategy for the impact analyses of the prospective data sets. Furthermore, such analyses can yield valuable information. When practiced skillfully and sensitively, such analyses can yield quite credible causal statements.⁴³ Estimates that are consistent across several data sets and are robust under alternative model specifications will be especially credible.

The MDRC and SPD data sets are the most promising for yielding credible findings. Both contain many before- and after-PRWORA observations, allowing for firmer estimates of pre-PRWORA trends to extrapolate into the post-PRWORA period and better estimates of the important outcomes of employment and earnings. MDRC’s data sets are particularly promising because the nonsurvey information provides for more detailed understanding of the four PRWORA local programs.

The Child Impact Waiver experiments. The descriptive information resulting from the waiver experiments likely will not be of much more than local interest. Whatever benefits they may yield will come from impact estimates. To the extent that the changes in the experiments made the experimental conditions equivalent to the TANF plans in those states and control-group conditions are maintained with reasonable fidelity, the experiments will provide the only estimates of TANF effects based on randomized designs that will be available in the near future. The waiver experiments’ findings therefore will be of considerable importance.

⁴³ Over the past decade, considerable thought has gone into how best to analyze nonexperimental data (Heckman and Robb 1986; Rosenbaum 1995).

2: Research on PRWORA: What Can Be Learned

What Will Not Be Learned from the Four Studies

As emphasized in this paper, a major limitation to before-and-after designs is their inability to distinguish between the effects of historical events or trends and the effects of the program being developed. Accordingly, differences between pre- and post-PRWORA conditions cannot be uniquely attributed to PRWORA. If TANF rolls are much smaller than AFDC rolls, it could be a result of welfare reform or the consequence of other trends occurring in the post-PRWORA period. For example, in the several years before PRWORA, welfare rolls were declining; that trend simply might persist into the post-PRWORA period. Similarly, inequality in the distribution of household income has been increasing over the past decade or so. An increase in income inequality post-PRWORA may simply be a continuation of that trend. In short, the findings of the three projects will not reveal much about the overall effects of PRWORA. Of course, this limitation does not apply to the augmented waiver experiments because they are based on randomized designs.

The decline in welfare rolls also has the effect of considerably reducing the number of welfare families in each of the surveys taken after TANF went into effect. NSAF-I completed interviews with about 2,700 families receiving welfare assistance; the numbers in individual states ranged from 83 in Alabama to 379 in Wisconsin and averaged 188 in the thirteen states.⁴⁴ Because the rolls have further declined since 1997, NSAF-II will have considerably smaller sample sizes for welfare families, perhaps as much as 25 percent smaller. The small sample sizes will restrict the ability of analysts to estimate the impact of welfare reform, especially subgroup differences at the level of individual states.

The four projects also have another important limitation: As planned now, none of the research extends much beyond the first few years of PRWORA. For many states, the period covered will be one in which each state will be developing its version of PRWORA—writing regulations, training staff, and disseminating knowledge of the new system to its clients. For many states, this process will take several years to complete. The “permanent” version of PRWORA may not appear, at the earliest, until beyond the research period. The NSAF and UC second surveys may reflect a stage in the evolution of PRWORA before the transition to the permanent PRWORA versions will have occurred. The waiver experiments end in the period 1998 through 2000. Accordingly, the versions of PRWORA being studied in the four projects may not look like the PRWORA that exists later in the first decade of the twenty-first century. This timing, however, may have some strengths. For example, if PRWORA has some important unwanted effects—say, a sharp rise in the number of children in abject poverty—it would be useful to know about that as quickly as possible so that countermeasures can be undertaken.

It also is important to keep in mind that research on the effects of PRWORA will not come to an end when the four research projects are completed. Indeed, the projects may be

⁴⁴ By design, the Wisconsin sample was larger than in other states.

2: Research on PRWORA: What Can Be Learned

extended: An NSAF-III for 2001 is now in the planning stage; SIPP samples may be enlarged to provide more detailed information; and the waiver experiments can be extended. New research projects may be funded to carry on when these three projects have ended. It is almost certain that interest in learning more about poverty, unemployment, and PRWORA will continue indefinitely into the future.

Although the research projects cover many important outcomes that might be affected by PRWORA, they will not be able to track how the changing conditions of the poor affect child abuse, substance abuse, or housing adequacy because none of the studies plan to measure them. Missing also are measures of social approval or disapproval as experienced from peers, kin, or the larger society. Of course, limitations exist on how many outcomes can be measured with the kinds of survey instruments used, and well-tested, reliable, measurement instruments do not exist for some alternative outcome measures. Furthermore, the research designers may have thought that such areas of behavior were unlikely to be affected by PRWORA. In any event, even if all the studies successfully overcome data collection problems, not all questions concerning the changes accompanying welfare reform will be answered.

Finally, an important limitation arises because a large part of the 1994–1999 decline in enrollment may be a result of “entry effects,” wherein fewer eligible families apply for benefits and, perhaps, fewer are able to successfully complete the enrollment process. Anecdotal evidence indicates that some state welfare agencies have made the application process difficult to complete, thereby discouraging prospective applicants. Whatever the cause for the decline, the entry effects preceding and accompanying PRWORA are producing important changes in welfare.

Entry effects can take several forms. In some states, welfare workers are authorized to provide emergency cash payments to people who apply for TANF benefits in exchange for becoming ineligible for TANF benefits for a specified period of time (for example, six months). The payments often are offered when it appears that the applicant may need immediate financial help, such as paying rent arrears or repairing a car, rather than long-term support. Some welfare departments have made applying for TANF a tedious and lengthy process, with the result that some applicants never complete the process. Of course, some eligible families simply may have come to believe that the new work requirements of TANF may be more trouble than the payments are worth and have not applied at all.

The research projects will have various difficulties providing data that directly bear on entry effects. Because SPD is a longitudinal study, however, it is better designed to study those effects. SPD can identify eligible families and determine the socioeconomic and employment characteristics of those who do not apply at various points in time in the pre- and post-PRWORA periods.⁴⁵ Urban Change and Assessing the New Federalism may be able to study entry effects

⁴⁵ In addition, SPD plans to ask about reasons for applying for welfare, reasons for not applying among eligible respondents, and reasons for denied applications.

2: Research on PRWORA: What Can Be Learned

indirectly by contrasting the composition of enrollees before and after PRWORA. Least able to provide information on entry effects are the five Child Impact Waiver experiments. The experimental and control groups were formed by randomly allocating people already enrolled at specific points in time and cannot provide information on how the composition of new enrollees subsequently may have changed.

Balancing the Books

The three prospective welfare reform research programs reviewed in this paper were driven by a concern for collecting the best possible information on how PRWORA would affect poor households. The constraints that had to be met in designing the studies were formidable. Everyone agreed that to launch randomized experiments after PRWORA started was simply out of the question. Of course, the waiver experiments could be used, although they could provide information on just a few states. Except for administrative data and extending SIPP, there was no way to collect extensive series of pre-PRWORA observations covering many states. Administrative data did not contain measures of such critically important outcomes as child well-being, and the task of pooling data from states with different data systems was formidable. SIPP had some child well-being measures, but they were not as extensive as desired. And so on.

In their design phases, each of the new projects countered its constraints in different ways. Taken as a group, it can be argued that the studies were about the best bundle of studies that could have been put together. To be sure, arguments can be made that each project could be improved, but the improvements would likely result in marginal benefits.

As the saying goes, hindsight is 20/20. Taken together, the three prospective projects will spend about \$200 million, and the extension of the five waiver experiments will cost more than \$12 million more. Given this level of funding, would other strategies have been more productive? For example, would it have been more productive to strengthen SPD, possibly by making intensive efforts initially to raise response rates, provide for interviews every four months (rather than annual post-PRWORA interviews), and allow for following all the households in the two cohorts? Alternatively, might it have been better to strengthen NSAF-I and -II by designing them as in-person surveys, which typically experience better response rates than telephone surveys? Would it have been better to provide funds to MDRC to expand its community studies to statewide studies? Perhaps any one of those alternatives would have led to richer data sets, but none of them would have made estimating the impact of PRWORA any easier. On balance, the current investment in three separate approaches, each with significant faults, may have been the optimal strategy, resulting in complementary data sets.

Whether extending the five waiver experiments was worthwhile requires considering different issues. In all five instances, extending the experiments took advantage of the already sunk costs of carrying the experiments through 1997. The investments in extensions are minor compared with the total costs of the experiments, and the potential returns are quite high in

2: Research on PRWORA: What Can Be Learned

comparison. That said, it is doubtful that extending the waiver experiment in Iowa made any sense because the experiment essentially ended around the time that the state PRWORA plans were implemented. The waiver experiments in Connecticut, Florida, Indiana, and Minnesota have much greater potential to provide useful information.

Welfare Reform Research in the Future

In the near term, it is important to support some of the ongoing research projects by providing additional resources to strengthen their contributions. In particular, SPD will not be very useful unless response rates can be materially improved. Accordingly, it is important that the funds the Bureau of the Census needs in order to raise the response rates to SPD be appropriated. It is heartening that some funds are already at hand, but the bulk of the funding is not yet forthcoming. If SPD can be materially improved, it will provide extremely useful information on the changes accompanying welfare reform.

The effort to raise response rates should be monitored carefully: If at some point it becomes clear that response rates will not be materially improved, then the effort ought to be discontinued and the unexpended funds put to some better purpose (see below). In addition, serious consideration ought to be given to actions ranging from releasing the data sets with strong warnings about their limitations to suppressing their release entirely.

The prospects for improving NSAF through investment of additional resources do not appear to be good. The Urban Institute and Westat have done as much as possible to compensate for NSAF's weaknesses. The low response rate to NSAF-I is troubling and is irremediable except through careful weighting. However successful the weighting scheme may be—and even if NSAF-II has achieved a more acceptable response rate—two surveys, before-and-after, are quite a weak design for estimating net effects. The major value of the Urban Institute studies will come from the detailed descriptive data for the thirteen states. It will be difficult to capture the diversity of state policies in a statistical model, given the many variations chosen by states and the frequency of their change.

MDRC's Urban Change project is not far enough along to make any judgment concerning its prospects. Especially critical will be its ability to analyze cohort experiences using administrative data. Assuming successful statistical modeling of quality data, UC will provide good estimates of effects within four important localities, supplemented by qualitative data on four local welfare systems. It is clearly too early to judge whether any steps can be taken in the short term to improve UC or, indeed, whether improvements will be needed.

The ongoing waiver experiments could prove to be valuable if strong efforts are made to ensure the fidelity of control-group conditions in each experiment. The danger is great that the control groups will be treated inappropriately and that the members of the control groups will not understand that they are not subject to state TANF rules. Maintaining the integrity of the control

2: Research on PRWORA: What Can Be Learned

groups means not only more effort on the part of the contractors but also the possibility that additional funds need to be given to support training of agency personnel and to provide for more frequent reminders to control group members of the special rules governing their welfare benefits. Unfortunately, given that the experiments have been underway for nearly four years since the start of TANF, it may now be too late to bolster their integrity.

It is also quite clear that when the four projects have been completed and analyzed, their findings will leave many critical questions unanswered. Almost certainly, PRWORA will be shown to be successful in meeting some of its goals in some of the states and failing to meet other goals in others. Questions will be raised about the effectiveness of time limits, family caps, income disregards, and other elements of the reform bundle. To answer those questions, further research will be needed. What form should such research take?

Perhaps the best strategy over the next decade or so would be to authorize and fund randomized experiments testing variations on the administrative and policy bundles. Those studies could be accomplished through state initiatives, but they are not likely to happen without federal funding. For example, questions about the effectiveness of family caps in reducing fertility might best be answered by conducting randomized experiments in which the experimental group is subject to family caps and the control group is under a no-family-caps condition (or vice versa). Randomized experiments can also be designed to observe the effects of varying the generosity of income disregards. Factorial experiments might be started to test the effects of various combinations of provisions that make up administrative and policy bundles.

This research strategy should lead to the accumulation of knowledge about how best to design welfare to achieve the dual objective of providing a safety net for the poor and facilitating entry into employment and higher income. A more expanded version of the current ACF strategy is proposed here. The current ACF experiments are not designed to unbundle TANF as much as to test proposed additions to the bundle. To understand how the bundles work, it is necessary to design experiments that vary such critical elements as earnings disregards, family caps, and time limits.

Some progress has been made toward implementing this strategy: ACF has funded three experiments and has issued an RFP for a fourth to be funded in 2000. All the experiments are designed to test measures aimed at improving TANF. For example, an experiment in Virginia will test the effectiveness of postemployment services in helping TANF clients retain their newly obtained employment. An evaluation planned in the future will involve four to ten states in MDRC-run experiments on measures aimed at employment retention and advancement in employment. The new experiments are patterned after the waiver strategy followed in the past decade or so of AFDC. They should lead to the accumulation of knowledge about how best to design welfare to achieve the dual objective of providing a safety net for the poor and facilitating entry into employment and higher income.

2: Research on PRWORA: What Can Be Learned

Studies also are needed to provide detailed descriptions of how the poor will fare under the welfare policy changes instigated by PRWORA and whatever other policy changes occur. We should be planning now how best to collect the data that will support an empirically based understanding of what is happening to the poor and what policy changes are likely to improve their condition.

When AFDC was a more or less uniform national program, national surveys such as the CPS or the SIPP may have served the purpose of monitoring the well-being of the poor. However, as discussed earlier, devolution has meant that state-level rather than national-level data are needed. An obvious move would be to enlarge sample sizes of existing ongoing national surveys to provide adequate state sample sizes. NSAF provides a good example in its selection of a small sample of critical states, an approach that the national surveys, including SPD, might want to emulate.

It is likely that in the end, SPD will not be very useful. Hence, serious consideration ought to be given to bolstering other ongoing large-scale surveys. In particular, it would be useful to augment SIPP and CPS by enlarging their sample sizes, especially bolstering their coverage of poor families. Ideally, I would like to see the sample sizes in at least the largest states increased enough to support state estimates.

Up to this point, the CPS has provided good monitoring data on the condition of the poor for the nation as a whole. Expanding the CPS sample to provide detailed data on a sample of states—and expanding CPS variables to include more information on how families with children are faring would be extremely useful. Additional efforts also should be made to address the problem of underreporting of welfare receipt. A parallel expansion of SIPP to conduct annual panel studies in a sample of states, especially in the ten to fifteen states that contain most of the poor, would be able to provide information on post-PRWORA changes in some detail. I recommend that the National Research Council of the National Academy of Sciences or a similar body examine the suitability of using SIPP for this purpose, paying special attention to attrition and nonresponse and their impact upon obtaining valid and reliable analyses.

Finally, a serious issue is how to promote responsible analyses of these data sets. Neither the research nor the policy communities will be content with only descriptive analyses. If Wisconsin poor families are better off (or worse off) in 2001 than they were in 1997 but California poor families show an opposite pattern, then some analysts certainly will try to discern whether the differences between the two states' versions of TANF are the source of the difference. To some extent, we can expect that competition among analysts will provide constructive criticism. In any event, those who release public data sets should warn potential users about the limitations of their data as well as provide full and detailed documentation about the data sets.

Appendix

Post-PRWORA Changes Made to HHS' Ongoing Experiments

Note: This appendix is freely adapted from a summary of the changes supplied by Howard Rolston.

The specific changes made to the conditions governing the experimental and control groups in the five states involved in the Child Impact Waiver experiments are described below:

Connecticut

Changes affecting only experimental-group cases.

- **Participation allowances.** Connecticut originally provided fixed participation allowances for work-related child care and transportation expenses. If actual expenses exceeded the allowance, the state would pay a supplement for the difference. Under the changes, the transportation supplement was retained, but the supplement for child care expenses has now been eliminated for families with only one child “in care.” In addition, a flat-rate special allowance has been put in place for participation in short-term activities, such as workshops or training sessions, in place of paying for child care and transportation costs.
- **Work participation exemption criteria.** Exemption for caring for a child under age 1 is no longer available to teen parents.
- **Time-limit extensions.** A prohibition against time-limit extensions was removed for when recipients were penalized for failure to comply with work requirements. In its place, the state required “individual performance contracts” establishing conditions recipients must meet to qualify for benefit extensions.
- **Penalties.** In place of progressive reductions in grant amounts for repetitive failure to cooperate with effort to verify eligibility or child support enforcement, a full-family sanction is now imposed for the first violation. However, the sanction can be lifted immediately by demonstrating cooperation. Sanctions for employment services violations or unjustified quits/fires remain progressive.

Changes affecting both experimental and control cases.

- **Expansions in eligibility.** Women are now eligible throughout pregnancy instead of just the last trimester.

2: Research on PRWORA: What Can Be Learned

- **Reduction in penalties.** There is no longer a penalty for transferring assets.
- **New TANF ineligibility criteria.** The following are no longer eligible for benefits:
 - Individuals convicted of drug-related felonies
 - Individuals fleeing felony prosecution or confinement
 - Individuals failing to participate in digital imaging identification procedure
 - Individuals convicted of fraudulently collecting public assistance
 - Noncitizens who have not applied for citizenship
 - Noncitizens who have not been state residents for at least six months.

Florida

Changes affecting both experimental and control cases.

- **Sanctions for noncompliance with work requirements.** The first instance of noncompliance is met with full-family sanction, but benefits are reinstated immediately upon compliance. The second instance is met with full-family sanction, and full benefits are restored after thirty days of compliance, although benefits may be continued for children under age 17. The third instance of noncompliance is met with full-family sanction, which can be restored after three months of compliance. After six consecutive months of compliance, the sanction counter is set to zero.

TANF differences from experimental-group conditions.

- **Time limits.** TANF includes a five-year lifetime time limit in addition to the time limits applying to experimental cases. The experimental time limits restrict benefits to twenty-four months in any sixty-month period, with some exceptions.
- **Asset limits.** The TANF asset limit is \$2,000, whereas the limit for the experimental group is \$5,000. TANF motor vehicle exclusion is \$8,500, but it is \$8,560 for experimental-group cases.
- **Work participation exemption.** TANF exempts adults caring for a child under 3 months old, whereas the experiment exempts those caring for a child under 6 months old.

Indiana

In May 1977 the experimental-group conditions were changed to match those in the state's new TANF program. The control-group conditions were not changed.

2: Research on PRWORA: What Can Be Learned

The changes made to the experimental group were as follows:

- **Time limits.** Months of benefit receipt were calculated cumulatively rather than consecutively for a total of twenty-four months. Months of ineligibility caused by program sanctions are considered as months of receipt. In addition, a recipient may earn an additional month of assistance for each consecutive six months of full-time employment.
- **Minor parents.** Minor parents and their children are required to live with specified relatives with the income and resources of the host relative considered in determining eligibility.
- **Fraud penalty.** A recipient convicted of fraud will become ineligible for assistance for twelve months following the first and second offense and become permanently ineligible upon conviction of a third offense.
- **School attendance requirement.** Excessive unexcused absences by a dependent child results in referral to a case manager for diagnosis and the development of a plan for remediation. Remediation failure may result in fiscal sanction affecting both caretaker and child.
- **Family benefit cap.** A monthly voucher equal to the cash increment for another child is available for children subject to the family cap.
- **Child care payment.** In place of receiving assistance equal to the family's benefit payment level before employment, employed recipients may opt to receive only a payment to cover child expenses.
- **Child support enforcement.** Failure to cooperate with paternity establishment will result in denial of benefits.
- **Personal responsibility agreement.** Each parent or caretaker must sign a personal responsibility agreement.
- **Penalties for illegal drug use.** Drug abusers will be referred for assessment and treatment. Failure to comply with treatment will result in a sanction of \$90 per month.
- **JOBS volunteers.** Exempt individuals who volunteer to participate in the JOBS program will be sanctioned if they do not attend or participate regularly.
- **Eligibility.** Recipient eligibility is determined using the Federal Poverty Guideline for family size, instead of the 185 percent-of-need standard.

2: Research on PRWORA: What Can Be Learned

- **Voluntary quits.** The needs of someone who voluntarily quits a job or reduces hours of work will be disregarded in determining a family's eligibility and assistance payment for six months.
- **Transitional child care.** Transitional child care is limited to twelve months during the period of the demonstration.

After the implementation of PRWORA, Indiana also established a new randomized experiment. The new experimental group experienced PRWORA TANF conditions, whereas the controls the pre-PRWORA AFDC conditions.

Iowa

Iowa terminated its experiment shortly after the implementation of its TANF plan. Because the experiment had been in place for three to three and a half years and the experimental conditions closely match the state's TANF plan, the experiments findings are arguably relevant to TANF. The differences between the experimental conditions and TANF are as follows:

- **Work transition period.** In the experiment, earnings of new workers are disregarded in the initial four months of employment. Under TANF this disregard was abolished.
- **Time limit.** No time limit existed in the experiment. TANF adopted a five-year time limit.

Minnesota

The experimental and control conditions in Minnesota's experiment were maintained. The only difference between the experimental conditions and TANF is that time limits are imposed on TANF participants.

2: Research on PRWORA: What Can Be Learned

References

- Bavier, Richard. 1998. "An Early Look at the Effects of Welfare Reform." Draft memorandum. Washington, D.C.: Office of Management and Budget.
- Brick, Pat Dean. 1999. "A Descriptive Review of Ten Major Surveys." Draft memorandum. Washington, D.C.: Urban Institute.
- Brim, Orville G., and David L. Featherman. 1997. "Surveying Midlife Development in the United States." Unpublished paper, Life Trends, Vero Beach, Fla.
- Camasso, Michael J., Carol Harvey, Radha Jagannathan, and Mark Killingworth. 1998a. *A Final Report on the Impact of New Jersey's Family Development Program: Experimental-Control Group Analysis*. Brunswick, N.J.: Rutgers University.
- Camasso, Michael J., Carol Harvey, Radha Jagannathan, and Mark Killingworth. 1998b. *A Final Report on the Impact of New Jersey's Family Development Program: Results from a Pre-Post Analysis of AFDC Case Heads from 1990-1996*. New Brunswick, N.J.: Rutgers University.
- Connecticut Department of Social Services. 1997. *Child Impact Studies Project*. Hartford: Connecticut Department of Social Services.
- Coulton, Claudia, and Nandita Verma. 1997. "Using Social Indicators to Study Change in Low-Income Neighborhoods Before and After the Implementation of Welfare Reform." MDRC project design paper. New York: Manpower Demonstration Research Corporation.
- Edin, Kathryn. 1997. "Urban Change Project Ethnographic Component." MDRC project design paper. New York: Manpower Demonstration Research Corporation.
- Edin, Kathryn, and Laura Lein. 1997. *Making Ends Meet: How Single Mothers Survive Welfare and Low-Wage Work*. New York: Russell Sage Foundation.
- Florida Department of Children and Families. 1997. *Child Impact Studies*. Tallahassee: Florida Department of Children and Families.
- Friedman, Joseph, and Daniel H. Weinberg. 1980. "Housing Consumption in an Experimental Housing Allowance Program: Issues of Self-Selection and Housing Requirements." Unpublished paper.
- Gallagher, L. Jerome, Megan Gallagher, Kevin Perese, Susan Schreiber, and Keith Watson. 1998. "One Year after Federal Welfare Reforms: a Description of State Temporary Assistance

2: Research on PRWORA: What Can Be Learned

for Needy Families (TANF) Decisions as of October 1997.” Occasional Paper no. 6. Washington, D.C.: Urban Institute.

General Accounting Office. 1998a. *Welfare Reform: HHS' Progress in Implementing Its Responsibilities*. Report no. HEHS-98-44. Washington, D.C.: General Accounting Office.

———. 1998b. *Welfare Reform: States Are Restructuring Programs to Reduce Welfare Dependency*. Report no. HEHS-98-109. Washington, D.C. General Accounting Office.

Groves et al. 1998. *Preliminary Report on Non-response Studies of NSAF*

Heckman, James, and Richard Robb. 1986. “Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes.” In *Drawing Inferences from Self-Selected Samples*, edited by H. Wainer. Berlin: Springer-Verlag.

Hollister, Robinson. 1997. “What Works, for Whom? Learning from the Experience of Welfare Reform.” *Poverty Research News* (Summer): 3–6.

Indiana Division of Family and Children. 1997. *Child Impact Studies: Operational Phase*. Indianapolis: Indiana Division of Family and Children.

Iowa Department of Human Services. 1997. *An Application for Continuing Funding for the Operational Phase: Iowa Child Impact Study*. Des Moines: Iowa Department of Human Services.

Manpower Demonstration Research Corporation. 1997. “The Institutional Study Design: Urban Change Project.” MDRC project design paper. New York: Manpower Demonstration Research Corporation.

Massey, James T., Dan O'Connor, and Karol Krotki. 1998. Response Rates in Random Digit Dialing (RDD) Telephone Surveys. *Proceedings of the Section on Survey Methods*. Washington, D.C.: American Statistical Association.

Michalopoulos, Charles, Hans Bos, Robert Lalonde, and Nandita Verman. 1997. “Assessing the Impact of Welfare Reform on Urban Communities: the Urban Change Project and Methodological Considerations.” MDRC project design paper. New York: Manpower Demonstration Research Corporation.

Minnesota Department of Human Services. 1997. *Child Impact Studies: Operational Phase*. St. Paul, Minn.: Minnesota Department of Human Services.

2: Research on PRWORA: What Can Be Learned

Quint, J. 1997. "Urban Change: Implementation Study Design". MDRC project design paper. New York: Manpower Demonstration Research Corporation.

Riccio, James A. 1998. *A Research Framework for Evaluating Jobs-plus, a Saturation and Place-based Employment Initiative for Public Housing Residents*. New York: Manpower Demonstration Research Corporation.

Robins, P. K., R.R. Spiegelman, S. Weiner, and J. G. Bell (eds.). 1980. *A Guaranteed Annual Income: Evidence From a Social Experiment*. New York: Academic Press.

Rossi, P.H., and K. Lyall. 1976. *Reforming Public Welfare*. New York: Russell Sage.

Rosenbaum, Paul R. 1995. *Observational Studies*. New York: Springer-Verlag.

Urban Institute. 1997. *National Survey of America's Families: Questionnaire*. Washington, D.C.: Urban Institute.

———. 1999a. *Survey Methods and Data Reliability*. NSAF Methodology Reports no. 1. Washington, D.C.: Urban Institute.

———. 1999b. *1997 NSAF Sample Design*. NSAF Methodology Reports no. 2. Washington, D.C.: Urban Institute.

———. 1999c. *1997 NSAF Snapshot Survey Weights*. NSAF Methodology Reports no. 3. Washington, D.C.: Urban Institute.

———. 1999d. *1997 NSAF Benchmarking Measures of Child and Family Well-being*. NSAF Methodology Reports no. 6. Washington, D.C.: Urban Institute.

———. 1999e. *Early Nonresponse Studies of the 1997 National Survey of America's Families*. NSAF Methodology Reports no. 7. Washington, D.C.: Urban Institute.

———. 1999f. *1997 NSAF Response Rates and Methods Evaluation*. NSAF Methodology Reports no. 8. Washington, D.C.: Urban Institute.

Urban Institute Policy and Benefits Staff. 1998. "Comparing the NSAF Results to the Current Population Survey." Internal memorandum. Washington, D.C.: Urban Institute Policy and Benefits Staff.

Weinberg, Daniel H., Vicki J. Huggins, Robert A. Kominski, and Charles T. Nelson. 1997. *A Survey of Program Dynamics for Evaluating Welfare Reform*. Paper presented at the Statistics Canada Symposium XIV.

2: Research on PRWORA: What Can Be Learned