

1

Introduction

*Douglas J. Besharov**

Between March 1994 and December 1999, welfare rolls fell an amazing 54 percent from their historic high of 5 million families. And that is the national figure. Nineteen states had declines of more than 60 percent.

How much of this decline is due to welfare reform? How much stems from other factors, such as the strong economy? And, more important, has welfare reform improved or worsened conditions for low-income families?

About a quarter billion dollars is being spent on studies and surveys specifically designed to answer these and other questions. This volume contains Peter H. Rossi's assessment of the four major research projects assessing welfare reform, which together account for more than half of the spending in this area. Rossi, a professor emeritus at the University of Massachusetts at Amherst, is one of the nation's foremost experts on program evaluation. This volume also contains the invited responses of the senior researchers for each of the projects.

Rossi's Assessment

In "Research on PRWROA: What Can Be Learned from Four Large-Scale Projects Currently Underway," Peter Rossi describes the four major research projects assessing welfare reform: (1) the Survey of Program Dynamics (SPD), (2) the National Survey of America's Families (NSAF), (3) the Project on Devolution and Urban Change, and (4) the Child Impact Waiver experiments (see box 1-1).

*Douglas J. Besharov is the Joseph J. and Violet Jacobs Scholar in Social Welfare Studies at the American Enterprise Institute and a professor at the University of Maryland's School of Public Affairs and director of its Welfare Reform Academy.

Rossi notes that response rate problems plague both the SPD and the NSAF. In the SPD, only 50 percent of the original sample responded in both 1998 and 1999.¹ “Especially worrisome is that higher percentages of low-income households stopped cooperating,” he says, concluding that “SPD will not be very useful unless response rates can be materially improved.” Even if efforts to reduce future attrition are successful, he fears that the rates will not be high enough to satisfy most researchers. In the first NSAF survey, response rates were only 65 percent for families with children and only 62 percent for families without children.² Rossi characterizes those rates as “average for well-run national telephone surveys,” but he warns that “‘average’ may not be good enough for surveys that are highly policy relevant.”³

Moreover, given the variation in state welfare regimes, Rossi says that “PRWORA cannot be evaluated as a national program . . . only state programs can be evaluated.” Consequently, many of the national surveys that have yielded so much of the information we have about welfare dynamics and other subjects are of limited usefulness. Much greater attention must be devoted to state, local, and even neighborhood data. (An obvious move, he notes, would be to enlarge sample sizes of existing ongoing national surveys to provide adequate state sample sizes.)

Although NSAF’s large samples for each of thirteen states also make it possible to examine how families in states with different TANF plans have fared, the decline in welfare rolls has considerably reduced the number of welfare families in the survey taken after TANF went into effect. The resultant small sample sizes of welfare recipients 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. State samples of welfare recipients are even smaller in the SPD.

The first survey in the MDRC Urban Change project achieved what Rossi termed a “very good” response rate of 79 percent. “The project should provide a rich—and textured—picture

¹The 1992–1993 SIPP began with 50,000 households. The 1997 SPD “Bridge Survey” was limited to the 38,000 households that completed all of their earlier SIPP interviews (about 73 percent of the original sample). The response rate to the Bridge Survey was 82 percent, yielding a sample of 30,125 interviewed households. As a result, the cumulative response rate at that point was just 59 percent. Budget constraints then forced the Census Bureau to reduce the sample to 18,500 in 1998. It then oversampled low-income households, using weights to try to maintain national representativeness. (This change does not affect the measured response rate.) The 1998 SPD suffered more attrition, resulting in a cumulative response rate of 50 percent. The total response rate remained at 50 percent in 1999, because the Census Bureau made special efforts, including incentive payments, to bring some nonrespondents back into the sample. The Census Bureau has asked Congress to appropriate funds for incentive payments to induce past nonrespondents to participate. Specifically, by offering an initial \$100 incentive payment plus a \$40 incentive for each subsequent year, the Bureau hopes to raise the response rate to 63 percent by 2002.

²These rates were achieved only after substantial additional efforts that may have introduced other methodological complications to the analysis.

³In their response to Rossi, the Urban Institute researchers argue that comparisons to other surveys should be based on an alternate weighted response rate which is slightly higher (about 70 percent).

regarding the characteristics and circumstances of the most vulnerable families” Unfortunately, because the study is limited to selected poor neighborhoods in four cities, the findings cannot be generalized to the broader welfare population nationally or even to other urban neighborhoods.

In addition to questions regarding the integrity of the data, Rossi cautions that none of the studies will be able to provide credible estimates of the effects of welfare reform—and therefore should not be used to make causal statements about the impact of welfare reform on the condition of low-income children. Before-and-after studies, he explains, cannot convincingly distinguish the effects of welfare reform from other changes occurring at the same time, such as improvements in the economy and expansions in other social programs, such as the Earned Income Tax Credit. Rossi suggests, however, that:

some attempts to estimate the net effects of PRWORA generally, or TANF specifically, may turn out to be 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 difference 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 Urban Change data sets as well as in the Child Impact Waiver experiments. However, such a convergence of findings across data sets is not likely to occur often.

Moreover, because each state has its own unique welfare program and because these programs are evolving over time, it will be difficult to estimate the general effects of welfare reform. Cross-state comparisons also are also problematic because of the difficulty in controlling for socioeconomic differences between the states and target populations, as well as the difficulty in characterizing state welfare reform policies.

Rossi also summarizes five randomized experiments of state welfare reform programs, comparing each state’s welfare reform plan with the old AFDC program. According to Rossi, “The importance of these continuing experiments is considerable.” He believes these experiments come closest to testing the impact of welfare reform, but he points out that they, too, suffer from problems such as a lack of generalizability, the failure to isolate control groups from some elements of welfare reform, and the inability to capture entry effects.

Rossi concludes by outlining a research agenda for the future. He believes that additional efforts should be made to bolster the SPD and NSAF response rates. Especially if efforts in that regard do not appear to be fruitful, he recommends that consideration be given to enhancing other surveys, especially CPS and SIPP. As Rossi writes:

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 at least the sample sizes in the largest states increased enough to support state estimates.

Rossi predicts, however, that “neither the research nor the policy communities will be content with only descriptive analyses” and says that one way to “promote responsible analyses” would be for those who “release public data sets [to] warn potential users about the limitations of their data. . . .”

Finally, he recommends that randomized experiments be designed to test the impact of specific welfare reform provisions.

Researcher Responses

Daniel Weinberg, chief of the Housing and Household Economic Statistics Division of the Census Bureau, and Stephanie Shipp, assistant division chief for labor force statistics and outreach, respond to Rossi’s comments about the SPD. In “Using the Survey of Program Dynamics for Welfare Reform Evaluation,” they acknowledge that the SPD cannot estimate net effects directly (because of the absence of a control group or other satisfactory “counterfactual”), but they argue that it can:

through modeling, decompose the impact of economic changes and welfare reform. Several studies have done this using pooled time-series, cross-sectional data The same models could be tested using longitudinal data from the SPD for the 1992–2001 period. Furthermore, even “gold-standard” random-assignment impact studies use modeling to account for differential attrition from treatment and control groups.

Weinberg and Shipp concede that the SPD’s response rates are low, but argue that they are comparable to those of other major longitudinal surveys, such as the Panel Study of Income Dynamics (PSID) and the National Longitudinal Survey of Youth (NLSY).⁴ They also

⁴Their benchmark for comparison, however, was the response rate after about a dozen interviews, which represents a period of about twelve years, since the initiation of the PSID and NLSY, but only about six years for the SPD. SIPP surveyed respondents three times a year, which was used as the basis for SPD.

Box 1–1. Welfare Reform Evaluative Studies

The National Survey of America’s Families (NSAF). Part of the Urban Institute’s Assessing the New Federalism project, NSAF is a nationwide survey of 44,000 households conducted first in 1997 and again in 1999, with plans for an additional survey in 2001 and, possibly, 2003. The New Federalism project is intended “to provide researchers resources to monitor and assess the consequences of program changes and fiscal developments.” Thus, the NSAF captures information on the economic, health, and social well-being of children, adults under age 65, and their families. Although the survey is national in scope, the focus is on thirteen states that together account for about half of the nation’s population. The 1997 survey was intended to be a snapshot of family life on the eve of major policy changes, most notably welfare reform, with subsequent surveys capturing the period after welfare reform.

The Survey of Program Dynamics (SPD). Conducted by the Census Bureau, SPD is a national longitudinal survey that was authorized by the 1996 welfare reform law to “evaluate the impact of welfare reform.” The SPD is an extension of the 1992 and 1993 panels of the Survey of Income and Program Participation (SIPP) and will follow the same households for ten years, from 1992 to 2001. It collects detailed information on the socioeconomic characteristics of participating households, including family composition, program participation, and employment and earnings. Because it is longitudinal, researchers will be able to examine transitions on and off welfare programs. In some years, the survey includes special questions related to child and adolescent well-being. The SPD started with 50,000 households in 1992–1993 (as SIPP), but attrition and budget restrictions have steadily eroded the sample.

The Project on Devolution and Urban Change acknowledge that response rates are lower for people with lower incomes (those of greatest interest to researchers evaluating welfare reform), but they describe the differences as “not enormous.” Moreover, they argue, the SPD data are representative of the population when compared with the Current Population Survey (CPS) on a range of socioeconomic variables. , conducted by the Manpower Demonstration Research Corporation (MDRC), is studying the aftermath of welfare reform in four large cities (Cleveland, Los Angeles, Miami, and Philadelphia). The project has several components, including ethnographic studies in poor neighborhoods, implementation studies examining how welfare agencies implemented the new law, and studies assessing the impact of welfare reform on local social agencies and businesses. In each city, 1,000 single mothers who received AFDC in 1995 were interviewed in 1998, covering such topics as their experience with welfare, their labor market participation, and the well-being of household members, especially children. The survey achieved what Rossi termed a “very good” response rate of 79 percent. A second survey of 1,000 single welfare mothers, covering the same topics, is scheduled for 2001. Finally, detailed administrative records covering welfare receipt and employment and earnings have been collected since 1992.

The Child Impact Waiver Experiments, funded by the U.S. Department of Health and Human Services (HHS), began before enactment of national welfare reform to test alternatives to the old AFDC program. States received waivers only for rigorous, random-assignment experiments to measure the impact of their new policies. Many waiver projects resemble TANF programs. HHS supported expansions of evaluations in five states (Connecticut, Florida, Indiana, Iowa, and Minnesota) to include additional measures of child well-being. The studies concentrate on families with children between ages 5 and 12.

I: Introduction

acknowledge that response rates are lower for people with lower incomes (those of greatest interest to researchers evaluating welfare reform), but they describe the differences as “not enormous.” Moreover, they argue, the SPD data are representative of the population when compared with the Current Population Survey (CPS) on a range of socioeconomic variables.

Weinberg and Shipp then describe the steps the Census Bureau is taking to address the response-rate problem. A 1998 experimental study using monetary incentives to bring nonrespondents back into the survey led the Bureau to conclude that a modest cash payment could be effective in reducing attrition. If additional funding is made available, the Bureau plans to interview the people who dropped out during the early periods of the survey and to offer financial incentives to participate in the survey, hoping to increase response rates to 60-63 percent by 2002. (The incentives would include an initial payment of \$100 in the first year and \$40 in subsequent years.)

In “Three Ways to Research PRWORA,” Kenneth Finegold, a visiting scholar at the Urban Institute, and Fritz Scheuren, senior fellow at the Institute, explain that the focus of the Institute’s “Assessing the New Federalism” project is much broader than welfare reform, and covers child care, child support, health insurance, child welfare, and foster care, as well as other topics. (Like the SPD, it also has detailed questions on child well-being.) According to Finegold and Scheuren, an added advantage of the NSAF is its large sample sizes in thirteen states, which will allow researchers to conduct more detailed state-level analyses.

Finegold and Scheuren then address several of the concerns Rossi raises. They argue, for example, that, although the NSAF’s response rates were originally reported to be 62 or 65 percent (as Rossi reports), it is also appropriate to use an alternative weighting methodology that would bring the response rate up to about 70 percent. They maintain that this response rate is superior to that of most other surveys similar in scale. They also describe several steps they have taken to compensate for this response rate problem and to minimize the bias associated with nonresponse.

Finegold and Scheuren agree that determining causality is difficult and that even randomized experiments are plagued by numerous problems. They therefore advocate supplementing experiments with survey analysis and microsimulation. They explain that the strengths and weaknesses are different for each approach and that, optimally, what is needed is to use all three approaches to evaluate PRWORA.

Charles Michalopoulos, a senior research associate at the Manpower Demonstration Research Corporation (MDRC), generally agrees with Rossi’s description and assessment of MDRC’s Project on Devolution and Urban Change, but takes issue with certain specifics. He agrees with Rossi that randomized experiments are ordinarily the best way to understand the effects of new policies and that “TANF cannot be evaluated as a national program; only state TANF programs can be evaluated.”

I: Introduction

Michalopoulos, however, disagrees with Rossi's conclusion that the nonexperimental studies, including the Urban Change project, cannot support very credible impact estimates. He believes that the Urban Change project will do better than Rossi suggests, because of the wealth of data it will analyze, derived from administrative records, periodic surveys, implementation research, ethnographic interviews, neighborhood indicators, and research on local institutions. Assuming successful statistical modeling, he believes that Urban Change will provide good estimates of the effects of welfare reform in four counties.

Michalopoulos also takes issue with Rossi's assertion that the findings cannot be generalized to the nation or even to other urban neighborhoods. He points out that, in 1997, more than 10 percent of all welfare recipients were in the four study counties, "so they may be as representative as any four large urban counties could be." Moreover, he observes, this criticism applies to randomized experiments as well. He concludes by noting that no research effort is perfect, including Urban Change, but that the project will help determine whether or not TANF has harmful effects on the poor.

Howard Rolston, director of Planning, Research and Evaluation at the Administration for Children and Families (ACF), U.S. Department of Health and Human Services (HHS), is responsible for monitoring the random assignment evaluations of the five waiver experiments Rossi reviews. Rolston largely agrees with Rossi's assessment of what will and will not be learned from these projects. He also echoes Rossi's support for random assignment experiments to test policy variations and notes that "enormous opportunities are available not only to *do*, but to *learn* while doing" (emphasis in original). He describes some of the efforts of states and localities, in collaboration with HHS, to design and evaluate policy options using experimental design, including job retention and advancement strategies, supported work, and alternative approaches for implementing child care subsidies.

In response to Rossi's concerns about maintaining the integrity of the experiments, Rolston shares new findings indicating that "participants in the experimental and control groups correctly understand their respective policy regimes." (He also describes how the failure to inform research subjects properly about the policies to which they were subject led to the premature termination of a sixth, unrelated, experiment in Arizona.) Although Rolston concedes that the experiments are not perfect, he concludes that they should be regarded as "conservative estimates of how particular policies compare with Aid to Families with Dependent Children (AFDC), and as a more accurate estimate of their effects in an environment of national welfare reform."

Conclusion

If one accepts Rossi's assessment, we will not be able to gauge welfare reform's impact on low-income children and families with reasonable precision. Put simply, no existing or planned study can provide a reliable assessment of welfare reform's impact because it is too late to construct a valid control or comparison group with which to measure the counterfactual (that is, what would have happened in the absence of welfare reform). The declines in welfare caseloads, for example,

have undoubtedly been spurred by the contemporaneously strong economy and dramatically increased aid to the working poor.

That does not, however, end the inquiry. *It remains essential to monitor the condition of low-income children and families under welfare reform, leaving aside the question of causation.* (This includes not just children and families that have left welfare, but also those that who have remained on welfare and those who have never been on welfare.) Putting in place mechanisms able to do so should be the immediate public policy goal. For, how else will we determine whether welfare and other policies for low-income families need to be changed.

The four research efforts described in this volume should be helpful in monitoring the well-being of low-income families, but each has particular weaknesses that limit its contribution. Both the SPD and NSAF have low response rates that undermine their utility. But the limitations go deeper. For example, existing measures of income seem truncated in their inability to parse out the household economy of cohabiting and coresident relationships. One limitation is our inability to know how much money is shared among the adults in the same household.

In addition, it is likely that whatever welfare reform's impact, it will not affect all low-income families equally. Some families will benefit, perhaps greatly, and some will be harmed—again, perhaps greatly. Thus, any analysis will have to be detailed enough to identify the winners and losers in the post-welfare reform world and to quantify their gains and losses (what evaluators call “subgroup analysis”). Moreover, given the variation in state welfare regimes and economies, a prime need is to improve and refine data at the state, local, and even neighborhood level.

Improving the quality of data may mean redirecting resources, as Rossi suggests. For example, the additional information provided by the SPD and NSAF—compared to with what is already available through the CPS, SIPP, and other surveys—does not seem worth the more than \$100 million that will be spent on them. Instead, it would be worthwhile to consider whether additional funding for both CPS and SIPP would result in more reliable and more complete data. (The longitudinal aspects of SIPP could also be enhanced.)

Improving data collection efforts will not be an easy or short-term effort. It is, however, the only approach that promises reliable and continuing data with which to monitor the well-being of low-income children and their families. For the long term, that will be the key challenge—and ingredient—of responsible policy making.