

# 4

## **The National Survey of America's Families**

*Kenneth Finegold and Fritz Scheuren\**

The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 can be considered the centerpiece of welfare reform. Peter H. Rossi's paper examines the contributions of current research projects to PRWORA evaluation, including the Urban Institute's Assessing the New Federalism (ANF) project and its National Survey of America's Families. Rossi seems to prefer social experiments to survey analysis as ways of researching PRWORA. Among surveys, he seems to like the Census Bureau's Survey of Income and Program Participation and its offspring, the Survey of Program Dynamics, better than NSAF.

We welcome Rossi's efforts, which can only result in clearer thinking about the problems of analyzing the recent transformations of American social policies, but we disagree with some of his comments about ANF and NSAF. In this response, we suggest reasons for greater enthusiasm about the contributions of survey-based research, and we show that several of the potential problems that lead Rossi to be skeptical about NSAF have minimal effects or have been effectively addressed through methods such as poststratification reweighting. We also argue for greater caution about the limitations of social experiments such as the Child Impact Waiver studies and for greater attention to the possibilities of researching PRWORA through microsimulation, an approach Rossi overlooks that can integrate and build on what is learned from experimental and survey research.

### **ANF and NSAF Are Broader Than PRWORA, and PRWORA Is More Than TANF**

The goals of ANF and NSAF are much broader than evaluating PRWORA. ANF is a multiyear, multidisciplinary project aimed at understanding the devolution of responsibility for health care, income security, job training, social services, and other policies and the effects of this devolution on the well-being of children and their families. PRWORA is the most important piece of devolutionary legislation, but it is not the only one; indeed, the process of devolution was underway before PRWORA's passage in August 1996.

---

\*Kenneth Finegold is associate professor of government at Eastern Washington University and a visiting scholar with the Urban Institute's Assessing the New Federalism Project; Fritz Scheuren is senior fellow at the Urban Institute and director of the institute's National Survey of America's Families. The authors thank Linda Giannarelli, David Merriman, Elaine K. Swift, Alan Weil, and Michael Wiseman for their helpful comments.

#### **4: The National Survey of America's Families**

PRWORA, moreover, did more than just replace the Aid to Families with Dependent Children entitlement with the Temporary Assistance for Needy Families block grant, which is Rossi's focus. PRWORA also made major changes to Medicaid and to the Food Stamp Program, two low-income programs with more participants and higher costs than AFDC or TANF.

The importance of PRWORA provisions other than TANF and of devolutionary policies other than PRWORA is particularly evident in health care, which has been a major focus for NSAF and ANF. A recent ANF brief, for example, used NSAF data to estimate the number of adults who could potentially receive Medicaid under the new parental eligibility rules included in PRWORA (Dubay, Kenney, and Zuckerman 2000). Another recent ANF brief used NSAF data to identify patterns of access to dental care for low-income children, a topic of concern for parents and policymakers that is only indirectly related to the replacement of AFDC by TANF (Kenney, Ko, and Ormond 2000).

An NSAF designed solely to assess TANF might have been very different. It could have had, for example, more narrowly focused questions and screening, which would in turn have yielded larger samples of TANF recipients. Even higher response rates might have been achieved as well.

Nonetheless, we think that the survey as designed and conducted is more important for research on PRWORA and TANF than Rossi's evaluation might suggest. Researchers inside and outside the Urban Institute are already using ANF and NSAF data to analyze the effects of PRWORA. Douglas J. Besharov and Peter Germanis (2000, p. 28), for example, say, "The best source of data about the families that have left welfare are surveys of former welfare recipients ('leaver studies') that have been conducted by various states and by the Urban Institute." They suggest that one surprising finding from this data—that almost half the leavers are not working regularly—"has profound implications for the economic and social condition of low-income families" (p. 29). Such analyses demonstrate the value of these new information sources better than any theoretical or methodological arguments we can make here.

#### **NSAF and Survey Analysis**

Because state discretion, arguably, is the central idea of devolution, national samples, such as the Current Population Survey (CPS), the SIPP, and the SPD, may not adequately capture state-by-state effects. And those effects can be considerable: States have new responsibilities for designing programs that provide cash assistance for families with children, child support, food stamps, health insurance coverage for children, child care, and education and training programs for low-income adults.

NSAF provides larger samples of poor and near-poor households, at both the national and state levels, than do CPS, SIPP, or SPD. Table 4–1 shows the achieved 1997 NSAF sample sizes

#### 4: The National Survey of America's Families

for all households and for households under 200 percent of the poverty threshold. Comparisons to the March 1997 CPS are also provided, illustrating one of the gaps the NSAF fills: The CPS sample sizes, especially for households under 200 percent of the poverty threshold, were for the most part too small to allow for reliable estimates of this group by state. SIPP and SPD sample sizes for the low-income population were even smaller on a state basis.

**Table 4-1. Comparison of 1997 NSAF and 1997 CPS Sample Sizes**

Site	All Households		Households Below 200 Percent of the Poverty Threshold	
	NSAF	CPS	NSAF	CPS
Alabama	2,553	561	1,276	191
California	2,543	3,904	1,224	1,480
Colorado	3,175	678	1,249	195
Florida	2,368	2,018	1,158	724
Massachusetts	3,238	979	1,114	265
Michigan	2,776	1,392	1,061	362
Minnesota	3,285	573	1,182	134
Mississippi	2,390	518	1,293	229
New Jersey	3,567	1,249	1,193	296
New York	2,632	2,825	1,222	1,008
Texas	2,452	2,350	1,295	913
Washington	3,393	566	1,337	177
Wisconsin	5,355	607	2,111	145
Balance of United States	4,716	23,687	2,086	7,428
Total	44,461	41,907	18,801	13,547

Source: Scheuren and Wang 1999.

The 1997 NSAF is a large national sample with sizable oversamples in thirteen states (collectively covering more than half the population of the United States). As table 1 displays, interviews for the 1997 NSAF were conducted in more than 44,000 households, yielding information on more than 100,000 people. Wisconsin was targeted for particularly intensive study, with separate large samples for Milwaukee and the balance of the state.

For the first round of the NSAF, data were obtained from February to November 1997. The survey asked an extensive battery of questions on the economic, health, and social characteristics of children, adults under age 65, and their families. By design, households under 200 percent of the federal poverty threshold were oversampled. Westat conducted the data collection for the Urban Institute and Child Trends, Inc. The 1999 round of the survey has about the same sample sizes by state as the 1997 round; later rounds are expected to be similar in scope.

The NSAF is a dual-frame survey with two separate components. One is a random-digit dialing (RDD) survey of households with telephones. The RDD approach is a cost-effective way to collect the desired data. However, because households without telephones (“nontelephone households”) contain a significant proportion of low-income children, a supplementary area

#### 4: The National Survey of America's Families

sample was conducted in person for those households. In the area sample, households within sampled blocks were screened, and all non-telephone households with someone under age 65 were interviewed. The dual-frame procedures pioneered for the 1997 NSAF were so successful that they have been repeated, with few modifications, in the 1999 survey.

All telephone interviewers worked in central interviewing facilities using computer-assisted telephone interviewing (CATI) technology. In-person interviewers provided cellular telephones to respondents in nontelephone households to connect the respondents to the interviewing centers for the CATI interview. Nontelephone household interviews were conducted in essentially the same way as for RDD households. At least 10 percent of each telephone interviewer's work was silently monitored for quality control purposes. For a more complete summary description of NSAF, see Scheuren and Wang (1999). Full details on the survey are to be found in the NSAF Methodology Series.<sup>1</sup>

Even if examined only from the point of evaluating PRWORA, NSAF does reasonably well in achieving appropriate sample sizes for the thirteen targeted states. The design of the NSAF sample, for example, means that it has many more families receiving AFDC/TANF than in the comparable (that is, March 1997) CPS (table 2). Despite the praise that Rossi has for SIPP/SPD, those surveys have many fewer AFDC/TANF recipients than either CPS or NSAF.

**Table 2. Families Receiving Public Assistance in the 1997 NSAF and 1997 CPS**

Site	Number of Families Receiving Public Assistance	
	NSAF	CPS
Alabama	103	20
California	271	274
Colorado	146	15
Florida	198	87
Massachusetts	269	43
Michigan	208	67
Minnesota	236	23
Mississippi	147	25
New Jersey	189	38
New York	215	255
Texas	169	67
Washington	279	29
Wisconsin	401	20
Balance of United States	331	924
Total	3,162	1,887

Source: Scheuren and Wang 1999.

**Criticisms of NSAF response rates.** Rossi's discussion of NSAF response rates is, at best, incomplete. NSAF response rates both are higher than he reports and compare favorably with those of other household surveys.

<sup>1</sup>See <http://newfederalism.urban.org/nsaf/methodology.html>

#### 4: The National Survey of America's Families

Two NSAF response rates<sup>2</sup> have been used, although others (for example, Scheuren 2000) may have value as vehicles for quantifying that survey's limitations. Rossi cites only the lowest of these, "65 percent for families with children and 62 percent for families without children," without drawing out the distinctions made elsewhere in the extensive NSAF Methodology Series. For comparisons with other surveys, we have been using a different weighted response rate, about 70 percent, for the 1997 NSAF, and we indicate in our findings that we expect the weighted rate to be only slightly less for the 1999 NSAF. The discrepancy arises because in the 1997 NSAF documentation, we give two response rate calculations. The operational rates (at 62 percent and 65 percent) originally were used and are documented in Brick, Flores-Cervantes, and Cantor (1999). The second rate calculated (at about 70 percent) was chosen to be more nearly comparable to that used in other surveys; it is described in Brick, Kenney, et al. (1999).

How do NSAF response rates at about 70 percent compare with other household surveys similar in scale? Such a comparison indicates that the NSAF response rates must be considered to be at the high end of the spectrum—certainly not average. The most recent published review of large national Random Digit Dialing (RDD) surveys is by Massey et al. (1997), who show that the median response rate for RDD surveys is below 60 percent. Less than 20 percent of the surveys they reviewed had response rates above 64 percent.

Household response rates for the CPS, perhaps the best-known non-RDD survey, are considerably better than in NSAF, at around 93 percent. One reason for the high rate is that the CPS starts out as a face-to-face survey and calculates its response rates differently. The NSAF area component offers the closest parallel; for this part of the NSAF, we achieved response rates of about 80 percent. Another important reason for the difference is that the CPS allows proxies: Any responsible adult may answer questions for the household. In the NSAF, designated respondents generally were required. The differences in respondent rules have the effect of trading off lower response rates (in the NSAF) for potentially larger measurement errors (in the CPS). This effect is probably not small; in fact, allowing proxies might have increased the NSAF response rate by up to 10 percent.

**Concerns about nonresponse bias.** No matter what the response rate is, survey estimates will be unbiased only when no differences exist between respondents and nonrespondents on survey items of interest. Thus, although nonresponse bias can increase as the response rate decreases, the response rate is not in itself a direct indicator of the magnitude of nonresponse bias.

Because of its importance in understanding devolution, great efforts were undertaken in NSAF to examine differences between the characteristics of NSAF respondents and nonrespondents; the work is documented in detail in Groves and Wissoker (1999). This was not the only effort made. An extensive set of comparisons was carried out between NSAF and several other studies with overlapping content (for example, Brick 1999), notably the CPS, the

---

<sup>2</sup>Ibid.

#### **4: The National Survey of America's Families**

National Health Interview Survey, and the SIPP, among others. A summary of those comparisons is forthcoming in Brennan et al. (2000).

The 1997 NSAF results just mentioned and the continuing work done on nonresponse for the 1999 NSAF (Scheuren 2000) have found no evidence of large or systematic nonresponse errors in the NSAF statistics examined. Even before poststratification adjustments for nonresponse, the special study we did for 1997 shows little, if any, bias. After adjustment, the remaining differences between NSAF and the results from comparable surveys (like CPS, which has better response rates) are minimal. In summary, the major results from NSAF, after adjustment, are robust against nonresponse bias.

**Analysis of NSAF.** Other sources of error in NSAF are similar to the problems of other large-scale household surveys, and NSAF arguably may be better off than most. In addition to nonresponse, measurement errors (for example, those caused by respondents misunderstanding certain questions) and, of course, sampling error (despite NSAF's large size) can affect research uses of the data. Researchers will need to bear in mind ordinary survey limitations when doing analyses. To this end, our releases of NSAF public use files, as Rossi recommends, "warn potential users about the limitations of their data as well as provide full and detailed documentation about the data sets."

The NSAF Methodology Series (now well along for both 1997 and 1999) is our attempt to do exactly what Rossi wants (see also Scheuren 2000, for an example). When finished, the 1997 series will have twenty-two volumes. We have already come quite far in developing a complete metadata system (for example, Dippo and Sundgren 2000) around the 1997 NSAF (with nineteen volumes completed to date). More than 2,000 researchers have registered as users of the existing public use files. Widespread availability will make for the needed "competition among analysts" (as Rossi puts it)—access that promises to speed up the search for new knowledge. The data are indeed complex, but the documentation and research reports that the Urban Institute and Child Trends are producing do help in warning about the data's limitations and how to work around them.

#### **Social Experimentation and the Child Impact Waiver Studies**

Researchers trying to assess PRWORA should be at least as cautious about using data from the Child Impact studies or future social experiments as they are about using data from NSAF or other surveys. Randomized experiments, in our experience, generally cannot be readily scaled up to look at overall effects, can have serious nonsampling problems (for example, inability to track affected individuals), and can be badly out of date when completed—testing policy options that are not actually those implemented.

Rossi describes randomized experimentation as the "'gold standard' design for estimating net effects" of policy change and expects the Child Impact studies of waiver experiments in

#### 4: The National Survey of America's Families

Connecticut, Florida, Indiana, Iowa, and Minnesota to produce more “credible” causal findings than the other, nonexperimental research projects. He further endorses randomized experimentation as the best strategy for future PRWORA research. Yet Rossi also identifies problems with the experimental approach and its application to the five Child Impact Waiver study states. In the following section, we discuss the serious issues he raises, along with several other issues. Some of the problems are intrinsic to or common with the experimental approach. A federal system creates additional problems in the conduct of social experiments.

**“Bundling.”** As Rossi notes, the experimental treatments in each of the waiver states combined several distinct welfare reform provisions. Only Minnesota followed a research design that permits estimation of their separate effects.

**Tracking participants.** It is difficult to track participants over time, particularly those who leave the location of the experiment. This attrition becomes especially important when the potential outcomes of the experiment include changes in the probability of outmigration, as the controversial “welfare magnet” hypothesis would suggest.

**“Saturation” effects.** Experiments affecting a small number of randomly selected participants do not allow assessment of the “community” or “saturation” effects that could follow from universal implementation (Greenberg and Shroder 1997). Many observers, for example, have tied the success of welfare reform to changes in agency culture that lead caseworkers to view their tasks in terms of job placement rather than benefit calculation. Widespread implementation of welfare reforms might affect recipients’ wages or rents in ways that experiments on a small group of recipients, whose identities are presumably unknown to employers or landlords, would not. These issues are sometimes addressed with quasi-experimental designs, in which the experimental treatment is applied to all participants at a particular site and experimental outcomes are compared with those at a matched control site, but then the closeness of the site match becomes a new source of concern.

**Differences between experimental treatments and actual policies.** Because it takes time to conduct a randomized social experiment, the treatments tested often are not those that turn out to be central to policy debates, or the key provisions of legislation are approved and implemented before the experimental data have been collected and analyzed. Rossi acknowledges differences between the welfare reforms tested under the Child Impact study waiver states and those contained in PRWORA, which lead him to question the usefulness of the Iowa data collected after that state implemented TANF. He concludes, however, that in the other four states, the correspondence between the welfare waivers and TANF is “close enough.” This judgment is too sanguine. The Appendix lists thirteen changes that were made to bring the Indiana experiment under PRWORA; although Rossi finds most of those changes “quite minor,” surely thirteen minor changes can add up to major discontinuities.

#### 4: The National Survey of America's Families

The only change Rossi lists for Minnesota is the implementation of the five-year lifetime limit on TANF benefits, but that may be the single most important provision in PRWORA. Early findings suggest that even before recipients become subject to loss of benefits, they take time limits into account and “bank” months of eligibility for use later, when recipients might have greater need for assistance or qualify for higher benefits (Grogger and Michalopoulos 1999). Of the eleven states that implemented some form of time limit before passage of PRWORA, only Delaware had a lifetime limit, and its provisions included more generous exemptions than PRWORA and had not yet been implemented statewide (Gallagher et al. 1998; Savner and Greenberg 1995). PRWORA thus contained stricter time limits than those implemented in any state under waivers; the Minnesota welfare waiver, in contrast, contained no time limits at all. This difference may help explain why caseloads have dropped under PRWORA in Minnesota, as in every other state, whereas caseloads actually increased under the Minnesota waiver experiment (Knox, Miller, and Gennetian 2000).

**Unpredictable costs.** The Minnesota waiver experiment also illustrates another problem of social experiments: Unless benefits are rationed (which would introduce new issues of selection bias and implementation procedures), it is not possible to fix costs in advance. At the beginning of an experiment, its effects on welfare receipt and work activity are unknown. Therefore, as long as members of the experimental group who are eligible for benefits can receive them, the costs of the experiment are also unknown and cannot be held equal to the costs with the control group. In the Minnesota experiment, outcomes generally were better than in the control group, but costs were higher, too. Thus, comparisons between the experimental and control populations do not directly indicate whether the more positive outcomes resulted from the experimental treatment per se or from the input of additional resources. Post hoc statistical adjustments or cost-benefit analysis can be applied to the data, but at best, either approach will approximate what the results would have been if costs had been constant.

**Federal diversity.** The diversity of the American federal system makes it difficult to know whether the experimental treatments would have similar effects in other states. President George Bush, arguing for federal approval of state waiver requests, said, “These states aren’t all the same. Welfare problems in Milwaukee are quite different than those in Juneau, Alaska, for example, or in California someplace” (see Teles 1998, 136). The obvious diversity that might justify policies that differ in Wisconsin, Alaska, and California, however, also might limit the “generalizability” of waiver data across those states.

**Nonrandom site selection.** At each site, participants in an experimental treatment are selected randomly to avoid selection bias, but the sites themselves are not chosen by random selection. Rather, sites are selected through a politicized waiver process in which state officials request and implement reforms, subject to approval by federal officials. To the extent that variables affecting waiver outcomes also determine which states seek waivers, how federal agencies respond to the requests, or how waiver provisions are implemented, waiver outcomes will yield biased predictions about outcomes from implementation in all states.



#### **4: The National Survey of America's Families**

Compared with the average state, for example, Minnesota has a more equal income distribution, lower unemployment, a more competitive two-party (and now, perhaps, three-party) system, a moralistic political culture, below-average proportions of racial and ethnic minorities and of recent immigrants, higher spending, and a more progressive tax system (Burt, Geen, and Duke 1997; Gray, Hanson, and Jacob 1999). It has also ranked consistently as the healthiest state in the nation (Coughlin et al. 1997). At least some of those variables are plausibly related both to Minnesota's decision to test a relatively liberal welfare reform package, which emphasized support and incentives over sanctions and did not include time limits, and to the positive outcomes of the experiment.

#### **Microsimulation as a Third Approach**

None of the above concerns should be taken to mean that experimental data cannot be valuable or that future experiments along the lines proposed by Rossi should not be pursued. Of course, they can be carried out only to the extent permitted under PRWORA, which, for example, prohibits states, even with waivers, from eliminating or weakening mandatory work requirements. Clearly, assessments of PRWORA based on experimental data will be most credible when they can be corroborated with data from surveys or microsimulation.

Microsimulation, which Rossi does not discuss as an approach to researching PRWORA, addresses some of the problems of social experiments or surveys. The Urban Institute's TRIM3 and other microsimulation models calculate the effects of complex, large-scale governmental tax, transfer, and health programs at the individual, family, state, and national levels (Giannarelli 1992; Citro and Hanushek 1991; Lewis and Michel 1990). Microsimulation models are based on data from surveys such as the CPS, but the survey data are transformed by adjustments to match administrative aggregates and imputations of missing information. The resulting data can be used to explore patterns of program eligibility and participation, interactions in the effects of different programs, and what-if experiments that alter program parameters such as benefit levels or eligibility criteria.

One advantage of microsimulation is that its estimates of the number of program participants and the amount of benefits they receive are aligned to administrative totals. This process adjusts for the underreporting of transfer payments, which, as Rossi notes, has been a consistent problem in the CPS and other surveys. By applying the rules of each complicated and interrelated program to each unit, microsimulation also generates consistent eligibility data, which is usually not available directly from survey or experimental sources. Only with microsimulation, therefore, is it possible to directly estimate the relative contributions of participation and eligibility trends to caseload changes such as the decline in TANF and food stamp recipients since 1996.

#### **4: The National Survey of America's Families**

Another advantage of microsimulation is that data for a long time series is already available. Assessments of PRWORA based on experimental or survey data often rely on the convenient fiction that control group data from waiver experiments or survey data that predate implementation of PRWORA measure conditions “before” welfare reform. In truth, welfare reform has been underway since the 1980s, as the Reagan, Bush, and Clinton administrations each encouraged states to request waivers. As the latest in a series of microsimulation models developed since the 1960s with support from the U.S. Department of Health and Human Services, TRIM3 can be used to track changes in participation and eligibility throughout that period. Microsimulation also makes it possible to estimate the impact of, say, 1997 TANF rules on 1995 CPS data, or of 1995 AFDC rules on 1997 data, thus controlling for some of the demographic changes that confound other efforts to estimate the effects of changes in program rules.

Yet another advantage of microsimulation is that it can be used to estimate the effects of potential future changes in program rules more quickly and more cheaply than can be done through social experiments. Microsimulation also avoids what social experiments cannot: the ethical issues and social consequences arising from the possibility that real human beings will be harmed as their TANF benefits, food stamps, or access to health insurance are altered. For these reasons, HHS frequently has used TRIM3 to test possible changes in AFDC and TANF, and the U.S. Department of Agriculture has used Mathematica’s microsimulation model of the Food Stamp Program for similar purposes.

One final advantage to mention is the flexibility that microsimulation models have in absorbing disparate data sources and behavioral insights. Many microsimulation datasets are amalgams of several data sources. Numerous data-handling issues arise around the creation of such amalgams (for example, Moriarity and Scheuren 2000a, 2000b), but when done carefully, microsimulation can represent an excellent heuristic for combining complex information (National Research Council 1992).

Microsimulation, of course, has limitations. TRIM3 and other “static” models apply new or experimental program rules to existing data, thus assuming that program changes do not generate such behavioral changes as increased or decreased employment or marriage rates. In contexts such as welfare reform, where behavioral changes are feasible or are themselves of great interest, microsimulation data is best viewed as indicating the limits to expected effects. To the extent that PRWORA encouraged more welfare recipients to work or change their living arrangements, for example, microsimulation of 1995 CPS data and 1997 TANF rules would underestimate the caseload decline. This effect occurs because some of the recipients who worked more or married would lose eligibility and others would be eligible for lower benefits, which would in turn reduce their probability of participating. Another issue is that adjusting survey data to align them to administrative totals is a tricky statistical proposition. It is also necessary to build algorithms that capture new policy provisions and to test and calibrate model results.

### The Inescapable Problem of Causality

If we are to draw any conclusions about what PRWORA has wrought, what it might do in the future, or what effects changes in its provisions might have, we must infer causality. We know from administrative data that welfare caseloads have declined, and we can obtain additional descriptive information from surveys such as NSAF or from the ethnographic studies that are part of the MDRC Devolution and Urban Change project discussed by Rossi. But those sources cannot directly tell us why changes have occurred. For example, caseload trends do not by themselves indicate the extent to which they were caused by welfare reform or by the unusually strong macroeconomic conditions of the past few years.

Rossi's evident hope is that data from the social experiments will make these inferences clear-cut. The Child Impact Waiver studies, however, demonstrate that inferences from social experiments to public policy are complicated by problems both in the construction of the experiments and in the correspondence between the experimental treatments and the policy changes that are eventually implemented.

Survey analysis and microsimulation offer alternative methods of inferring causality. Questions of causality can be explored by comparing survey data from before and after a policy intervention, using appropriate statistical techniques to control for other independent variables. Conclusions derived from survey analysis become even more credible when time-series data is also cross-sectional, as is true for NSAF; pooled designs then can be used in sophisticated tests of alternative causal hypotheses. Microsimulation can be used to explore causality by running preintervention rules on postintervention data (or vice versa) and by conducting sensitivity analyses that test the impact of changes in programs on outcomes of substantive interest. Data from the first (1997) and second (1999) rounds of NSAF are already publicly available. TRIM3 data for 1997 will be publicly available soon, enhancing the ability of researchers to grapple with the problems of causality that surround PRWORA.

### Summary

In this response, we have contrasted what can be learned from social experiments with what can be learned from survey-based research and from microsimulation, and we have argued for the need to use all three approaches to evaluate PRWORA. We believe that this view is more balanced than the conclusions presented by Rossi. What remains to be said is that all three approaches should be used together whenever possible: real experiments (for example, the Child Impact Waiver studies) and thought experiments (à la TRIM3 microsimulation) applied to representative populations of potential and actual recipients obtained through surveys (such as NSAF). The strengths and weaknesses differ for each approach.<sup>3</sup> Optimally, all three methods

---

<sup>3</sup> One example of how microsimulation modeling can borrow strength from a survey can be found in Ingram et al. (2000). The authors of that paper use NSAF data to test a crucial data-handling issue that

#### 4: The National Survey of America's Families

should be used to evaluate PRWORA. This is exactly what the Urban Institute and Child Trends are doing as part of the ANF project.

In a world in which public officials tested reforms one at a time and waited until the tests were completed and their results fully analyzed before approving any new legislation, randomized social experiments might give us all the information we needed to assess major policy changes. In the very different world of welfare reform, however, social experiments, surveys, and microsimulation all are necessary, and so is careful inference from the imperfect measures that each approach to policy research can provide.

#### References

- Besharov, Douglas J., and Peter Germanis. 2000. "Welfare Reform—Four Years Later." *Public Interest* 140 (Summer):17–35.
- Brennan, Niall, Leticia Fernandez, Joyce Morton, Jeffrey Passel, Adam Safir, Fritz Scheuren, Kevin Wang, Sheila Zedlewski, and Stephen Zuckerman. 2000. *1997 NSAF National Benchmarking Measures*. NSAF Methodology Reports no. 15. Washington, D.C.: Urban Institute.
- Brick, J. Michael, Ismael Flores-Cervantes, and David Cantor. 1999. *1997 NSAF Response Rates and Methods Evaluation*. NSAF Methodology Reports no. 8. Washington, D.C.: Urban Institute. Available from: [http://newfederalism.urban.org/nsaf/methodology\\_rpts/Methodology\\_8.pdf](http://newfederalism.urban.org/nsaf/methodology_rpts/Methodology_8.pdf)
- Brick, Pat Dean. 1999. "A Descriptive Review of Ten Major Surveys." NSAF Working Paper. Washington, D.C.: Urban Institute.
- Brick, Pat Dean, Genevieve Kenney, Robin McCullough-Harlin, Shruti Rajan, Fritz Scheuren, Kevin Wang, J. Michael Brick, and Pat Cunningham. 1999. *Survey Methods and Data Reliability*. NSAF Methodology Reports no. 1. Washington, D.C.: Urban Institute. Available from: [http://newfederalism.urban.org/nsaf/methodology\\_rpts/Methodology\\_1.pdf](http://newfederalism.urban.org/nsaf/methodology_rpts/Methodology_1.pdf)
- Burt, Martha R., Rob Geen, and Amy-Ellen Duke. 1997. *Income Support and Social Services for Low-income People in Minnesota*. Assessing the New Federalism State Reports. Washington, D.C.: Urban Institute. Available from: <http://newfederalism.urban.org/html/MNincome1.html>
- Citro, Constance F., and Eric A. Hanushek, eds. 1991. *Improving Information for Social Policy Decisions: the Uses of Microsimulation Modeling*. 2 vols. Washington, D.C.: National Academy Press.

---

can arise in building a microsimulation model that employs two or more survey data sets that have been statistically matched together. NSAF itself was introduced into the TRIM3 modeling system this year.

#### 4: The National Survey of America's Families

Coughlin, Teresa A., Shruti Rajan, Stephen Zuckerman, and Jill A. Marsteller. 1997. *Health Policy for Low-income People in Minnesota*. Assessing the New Federalism State Reports. Washington, D.C.: Urban Institute. Available from: <http://newfederalism.urban.org/html/MNhealth1.html>

Dippo, Cathryn, and Bo Sundgren. 2000. "The Role of Metadata in Statistics." Paper presented at the American Statistical Association's International Conference on Establishment Surveys—II, Buffalo, N.Y., June.

Dubay, Lisa, Genevieve Kenney, and Stephen Zuckerman. 2000. *Extending Medicaid to Parents: an Incremental Strategy for Reducing the Number of Uninsured*. Assessing the New Federalism Policy Brief no. B-20. Washington, D.C.: Urban Institute. Available from: [http://newfederalism.urban.org/html/series\\_b/b20/b20.html](http://newfederalism.urban.org/html/series_b/b20/b20.html)

Gallagher, L. Jerome, Megan Gallagher, Kevin Perese, Susan Schreiber, and Keith Watson. 1998. "One Year after Federal Welfare Reform: a Description of State Temporary Assistance for Needy Families (TANF) Decisions as of October 1997." Assessing the New Federalism Occasional Paper no. 6. Washington, D.C.: Urban Institute. Available from: <http://newfederalism.urban.org/pdf/oneyr.pdf>

Giannarelli, Linda. 1992. *An Analyst's Guide to TRIM2: The Transfer Income Model, Version 2*. Washington, D.C.: Urban Institute Press.

Gray, Virginia, Russell L. Hanson, and Herbert Jacob. 1999. *Politics in the American States: a Comparative Analysis*. 7th ed. Washington, D.C.: CQ Press.

Greenberg, D. and M. Shroder. 1997. *The Digest of Social Experiments*. 2d ed. Washington, D.C.: Urban Institute Press.

Grogger, Jeff, and Charles Michalopoulos. 1999. "Welfare Dynamics under Term Limits." NBER Working Paper no. 7353. Cambridge, Mass.: National Bureau of Economic Research.

Groves, Robert, and Douglas Wissoker. 1999. *Early Nonresponse Studies of the 1997 National Survey of America's Families*. NSAF Methodology Reports no. 7. Washington, D.C.: Urban Institute. Available from: [http://newfederalism.urban.org/nsaf/methodology\\_rpts/Methodology\\_7.pdf](http://newfederalism.urban.org/nsaf/methodology_rpts/Methodology_7.pdf)

Ingram, Dedun, John O'Hare, Fritz Scheuren, and Joan Turek. 2000. "Exploiting Coincidences in Statistical Matching." *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, Va.: American Statistical Association, forthcoming.

Kenney, Genevieve M., Grace Ko, and Barbara A. Ormond. 2000. *Gaps in Prevention and Treatment: Dental Care for Low-Income Children*. Assessing the New Federalism Policy Brief no. B-15. Washington, D.C.: The Urban Institute. Available from: [http://newfederalism.urban.org/html/series\\_b/b15/b15.html](http://newfederalism.urban.org/html/series_b/b15/b15.html)

#### 4: The National Survey of America's Families

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

Lewis, Gordon H., and Richard C. Michel, eds. 1990. *Microsimulation Techniques for Tax and Transfer Analysis*. Washington, D.C.: Urban Institute Press.

Massey, James, Dan O'Conner, and Karol Krotki. 1997. "Response Rates in Random Digit Dialing (RDD) Telephone Surveys." *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, Va.: American Statistical Association, forthcoming.

Moriarity, Chris, and Fritz Scheuren. 2000a. "Statistical Matching: a Paradigm for Assessing the Uncertainty in the Procedure." *Journal of Official Statistics*, forthcoming.

Moriarity, Chris, and Fritz Scheuren. 2000b. A Note on Rubin's "Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations." *Journal of Business and Economic Statistics*, under review.

National Research Council. 1992. *Combining Information: Statistical Issues and Opportunities for Research*. Washington, D.C.: National Academy Press.

Savner, Steve, and Mark Greenberg. 1995. *The CLASP Guide to Welfare Waivers: 1992-1995*. Washington, D.C.: Center for Law and Social Policy.

Scheuren, F. 2000. "Quality Assessment of Quality Assessments: New Tools for a New Focus?" *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, Va.: American Statistical Association, forthcoming.

Scheuren, Fritz, and Kevin Wang. 1999. "The National Survey of America's Families: a Mid-Course Review." Paper presented at the American Enterprise Institute Conference on Child Well-Being Under Welfare Reform, Washington, D.C., December. Available from: [http://newfederalism.urban.org/nsaf/methodology\\_rpts/Methodology\\_16.pdf](http://newfederalism.urban.org/nsaf/methodology_rpts/Methodology_16.pdf)

Teles, Steven M. 1998. *Whose Welfare? AFDC and Elite Politics*. Lawrence: University Press of Kansas.