New Federalism and Research: Rearranging Old Methods to Study New Social Policies in the States

Stephen H. Bell

99–08

September 1999

Assessing the New Federalism
An Urban Institute Program to Assess Changing Social Policies
New Federalism and Research: Rearranging Old Methods to Study New Social Policies in the States

Stephen H. Bell
*The Urban Institute*
99–08
Assessing the New Federalism

Assessing the New Federalism is a multiyear Urban Institute project designed to analyze the devolution of responsibility for social programs from the federal government to the states. It focuses primarily on health care, income security, employment and training programs, and social services. Researchers monitor program changes and fiscal developments. Alan Weil is the project director. In collaboration with Child Trends, the project studies changes in family well-being. The project provides timely, nonpartisan information to inform public debate and to help state and local decisionmakers carry out their new responsibilities more effectively.

Key components of the project include a household survey, studies of policies in 13 states, and a database with information on all states and the District of Columbia. Publications and the database are available free of charge on the Urban Institute's Web site: http://www.urban.org. This paper is one in a series of discussion papers analyzing information from these and other sources.


The nonpartisan Urban Institute publishes studies, reports, and books on timely topics worthy of public consideration. The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, its funders, or other authors in the series.

Publisher: The Urban Institute, 2100 M Street, N.W., Washington, D.C. 20037
Copyright © 1999.
Permission is granted for reproduction of this document, with attribution to the Urban Institute.

The author is grateful to Henry Brady, Norman Bradburn, Charles Manski, and seminar participants at the Joint Center for Poverty Research at Northwestern University/the University of Chicago for comments on earlier presentations of this material, and to Doug Wissoker and Alan Weil for comments on the draft.
Abstract

As responsibility for social programs shifts more toward state and local government, the potential for path-breaking research on new policies is enormous. The Assessing the New Federalism (ANF) project at the Urban Institute examines recent social policy changes in the income security, employment, health, and social services areas and provides an example of how traditional evaluation methods can be adapted to address comprehensive state reforms.

As a first step, ANF’s research goals and structure are traced to shifting social policy responsibilities between the federal and state/local governments. Its state policy case studies concentrate on 13 states (AL, CA, CO, FL, MA, MI, MN, MS, NJ, NY, TX, WA, WI) that, as a group, are home to more than 50 percent of the nation's population. Similarly, ANF’s National Survey of America's Families collects state-reliable data on child and family outcomes for the same 13 states, and for the nation as a whole. Other consequences of devolution also generate new research challenges: many social policies changing at once; a lack of basic information on program activities and costs; and the shift from federal to state and foundation oversight of evaluation research.

In this context, ANF may be thought of as an early prototype for policy research in a devolved age. It has taught several lessons already—that (1) broad-based analyses are hard; (2) limits on research breadth and depth are essential; (3) data needs cannot be compromised (even if that means narrowing the analysis agenda); and (4) clear, timely presentation of results is more important than ever. Finally, the paper focuses on the largest analytic challenge of all, determining which state reforms are helping low-income citizens and advancing broader social objectives and which are not. Several issues and options are explored around this daunting problem, but not resolved.
# Table of Contents

New Federalism and Research: Rearranging Old Methods to Study New Social Policies in the States ................................................................. 1

Goals and Components of the *Assessing the New Federalism* Project ................................................................. 2

The Defining Features of Past Evaluations ................................................................. 5

Changes in the Evaluation Landscape with the Advent of the "New Federalism" ............ 8
  State-Specific Results ....................................................................................... 8
  Multiple Policy Changes ............................................................................... 9
  New Research Priorities .............................................................................. 12
  New Funding Mix ....................................................................................... 13

*ANF* Features That Reflect the Changing Landscape ........................................ 15

Lessons Learned to Date from the *ANF* Research ........................................ 16
  Broad-Based Analyses Are Hard ............................................................... 16
  Limits Are Essential ............................................................................... 19
  Data Needs Cannot Be Compromised ...................................................... 20
  Clear, Timely Presentation of Results Is More Important Than Ever ............ 22

The Final Challenge: Linking Policy to Outcomes ........................................ 23
  The Challenge ......................................................................................... 24
  Possible Responses ............................................................................... 26
  Next Steps .......................................................................................... 30

Notes ............................................................................................................. 32

References ..................................................................................................... 35

About the Author .......................................................................................... 36
New Federalism and Research: Rearranging Old Methods to Study New Social Policies in the States

Social programs in the areas of welfare, employment, health, and social services have been changing rapidly in the 1990s, especially since the passage of national welfare reform legislation (in 1996) and the children's health insurance initiative (in 1997). As responsibility for social programs shifts further toward state and local government, the range of policy directions that states may take—and, thus, the potential for path-breaking research in these areas—is enormous. The most direct transfer of authority stipulated by the "new federalism" movement is from federal to state governments. Hence, it is natural to begin research at the state level.¹

Assessing the New Federalism (ANF), an Urban Institute program that examines recent social policy changes, represents one of many state-focused studies emerging from the new policy environment. Thus, it provides a case study of how research might be designed to study comprehensive state reforms in general. The large number of policy questions the program seeks to address—and the vast, diverse "canvas" on which the answers are unfolding—necessitates the use of new, and potentially controversial, combinations of research and analysis tools. This paper outlines the design elements included in the ANF study and considers whether the research techniques used are likely to form a vital part of other full-spectrum analyses of comprehensive state policy reforms in the social welfare area.

The paper begins by describing ANF's research goals and their role in determining the project's design. Differences between this design and more familiar strategies for evaluating large-scale social programs are considered next and are traced to shifting social policy responsibilities between the federal and state/local governments. The remaining sections of the paper attempt to generalize from the ANF experience by considering some of the more global
challenges faced and lessons learned in conducting this type of research. Special attention is paid to the largest analytic challenge of all: determining which state reforms benefit low-income citizens—and advance broader social objectives—the most, all other things equal.

**Goals and Components of the Assessing the New Federalism Project**

*ANF* began in 1996 at the initiative of The Annie E. Casey Foundation, with major additional funding from the W. K. Kellogg Foundation, The Henry J. Kaiser Family Foundation, and The John D. and Catherine T. MacArthur Foundation. A joint effort by the Urban Institute, Child Trends, and Westat, *ANF* is both a large-scale research project and a vehicle for collecting data for use by researchers across the country. More than half of the effort will be used to produce data for public use, including:

- A **state database** of annual economic, demographic, and social indicators;
- The **National Survey of America's Families**, collecting a range of measures for 96,000 families and single adults (48,000 per wave); and
- Detailed **case studies** of programmatic and fiscal developments in 13 states, which contain more than 50 percent of the nation's population.

As a research project, *ANF* seeks to answer several broad questions:

- How much devolution in social policy is actually occurring—in dollars, program responsibilities, and "strings attached" by the federal government? What, if anything, is devolving further, from state to local governments?
- What policies are states adopting, and how are those choices being implemented? How do options for low-income families change as a result?
• Are desired changes in behavior—as defined both in Washington and in the states—taking place, including greater work effort, fewer out-of-wedlock births, and more stable families?

• Do these combined policy and behavioral changes lead low-income families and children to greater financial independence and well-being? To more secure medical coverage, child care arrangements, and home environments?

• Can general lessons be drawn about which state approaches are working best to achieve these outcomes, and which are producing unintended negative consequences?

• Are certain groups faring better or worse than others: families with multiple problems, the working poor, legal immigrants, the medically uninsured, and so on?

• How does the picture change if states engage in a "race to the bottom" or the nation encounters an economic downturn?

Consistent with this focus, the National Survey of America's Families will collect state-reliable data on key facets of family and individual well-being that may be affected by social policy. (Current national surveys have sufficient sample size to do this in only a few very large states.) Representative samples of families and individuals will be drawn in each of the 13 states included in the case studies (AL, CA, CO, FL, MA, MI, MN, MS, NJ, NY, TX, WA, WI) and for the balance of the nation. The survey is also distinguished by its wide-ranging questionnaire, which covers behaviors and outcomes in all of the areas potentially affected by devolving social policies:

• Health;

• Employment and income;

• Family composition;
• Housing, social services, and child care;
• Public assistance receipt; and
• Child and adult well-being.

The survey oversamples the 13 states, the low-income population, and families with children, yet can still provide a balanced cross-section of the nonelderly population of the United States as a whole. Data from the 1997 wave of interviews will be available to the public in 1999; a second cross-section of 48,000 households is planned for 1999 (with release to the public in 2000 or 2001).

The state database will draw largely on existing data sources to provide year-by-year indicators of states' situations from 1993 to 2000, including political, fiscal, economic, demographic, and policy conditions in each state and measures of program participation and family outcomes in the health, social service, income support, and employment realms. A detailed taxonomy of states' welfare policies, coded for analytic use, will also be included. These data are available for public use and can be downloaded from the Urban Institute home page at http://www.urban.org.

Detailed descriptive information on program administration in the 13 intensively studied states has been published in a series of income security/social service and health policy case studies. Reports on each state describe policy goals, program designs, and ground-level implementation based on in-depth site visits. Topical reports that cut across states provide additional synthesis and comparisons. Results from a second round of visits are expected in 2000–01. State fiscal analyses will be provided throughout the project.
The Defining Features of Past Evaluations

*ANF*'s design responds in a number of ways to the shift in program responsibilities from the federal level to state and local government. As we will explore in the remaining sections of the paper, the *ANF* design illustrates how devolution's programmatic shifts can lead to new research strategies, providing lessons for other studies of state-based social reforms.

To see how the design of *ANF* responds to devolution, we must first identify the features of federally sponsored programs that, in the past, have led to quite different designs for large-scale policy research. This discussion relies heavily on the extensive evaluation literature about federally sponsored employment and training programs.\(^2\) Thus, conclusions do not apply automatically to federal evaluations of other types of social programs (e.g., child care, Medicaid),\(^3\) though similar designs have been applied to policy research on those topics.\(^4\)

In the 1980s and 1990s, four aspects of federally initiated employment and training evaluations are worth noting:

- The interventions to be assessed were centrally defined, at least conceptually if not also operationally.
- For the most part, the program changes to be studied occurred incrementally and for one program at a time.
- With full accountability for results resting in Washington, federal agencies made tracking federal expenditures a continual high priority.
- Program evaluations were undertaken by organizations—federal agencies—with relatively large budgets and a relatively sophisticated understanding of research methods and their limitations.
Each of these contextual factors had a strong influence on how researchers conducted employment and training evaluations—and many evaluations in other fields—in the 1980s and 1990s.

With centrally defined interventions, a single national study was a reasonable guide to policy for all parts of the country. As a result, the sites and household samples included in the evaluations did not have to represent individual states or localities, as long as they reasonably characterized the country as a whole. This allowed researchers to gather the information they needed from a relatively small number of states (5 to 15, typically) and to confine micro-level data sets to a relatively small number of households or individuals per state. These data were adequate for national conclusions but not for site- or state-specific results. Other than special demonstration projects (i.e., small-scale pilot projects testing new program strategies), past national studies took relatively little interest in the individual program variations that tend to occur naturally among different geographic areas. The typical evaluation of the 1980s and 1990s tended to say nothing at all about individual states; evaluations did examine individual sites in many instances, but those sites were usually small areas (e.g., a single city or metropolitan area), not whole states.

Unitary, incremental program changes allowed researchers to clearly and narrowly define the intervention of interest. This definition facilitated the construction of good "counterfactuals" to the new policy component—representations of what the world would have looked like had a given policy change (in this case, a small, discrete policy change) not taken place. With sharply defined interventions, researchers could (through random assignment of individuals or other means) equalize the actual and counterfactual cases on all factors except the
policy change of interest. Thus, not only could researchers sharply define the intervention under study, they also had the tools to isolate its impact independent of other factors, including changes in other policies.

**National tracking of federal expenditures** also facilitated documentation and understanding of the intervention. Historically, sponsoring agencies continuously measured the fiscal—and many of the operational—traits of federally sponsored programs: dollars spent, number of participants, types of assistance provided, and so on. After years of tracking fairly stable programs, this gave the central government a good "fix" on what the nation was doing programmatically in a particular policy area. Also, when changes were made, the character and timing of those changes were self-evident to federal agencies, which were responsible for scheduling and orchestrating each of these policy reforms nationwide.

With strong program information coming in regularly, federal agencies could concentrate their research funds on collecting and analyzing outcome data on program participants. This meant that high-quality, standardized participant outcome data were collected for the nation on a regular basis. As a result, good national data were routinely available for not just outcomes but all aspects of federal programs: their cost, character, scale, timing, and results. This made the evaluator's job much easier and made research results more definitive at a relatively low marginal cost for data collection.

Finally, studies were usually undertaken by knowledgeable, well-funded sponsors. The involvement of technically savvy government agencies in the conception and direction of evaluations helped keep the topical scope of these studies under control and limit the agenda to research questions that were "answerable" using existing methodologies. That a single agency
completed almost all evaluations reinforced the focus on fairly narrow and technically manageable objectives. Containing a study's ambitions in these respects can be crucial to an evaluation's technical success and, thus, its credibility and influence within the policy community.

Changes in the Evaluation Landscape with the Advent of the "New Federalism"

Much of what was just described as the context and defining features of the "federal era" of social policy evaluations is now changing—at least as concerns anti-poverty policy—with the advent of state and local leadership. Particularly telling, in terms of the policy research strategies required by this new environment, are the following changes:

- Policy findings are now needed at the state level, where program authority and accountability now rest, rather than at the federal level.
- Whole sets of social policies are changing at once.
- The top-priority research questions have devolved to a more fundamental level, concentrating on basics such as "What programs are states and localities running?" "What services are they providing?" "Whom are they serving?" and "How much do these services cost?"
- The composition of research funders has shifted from federal agencies to state government and private foundations.

State-Specific Results

The need for findings at the state level is perhaps the most obvious and consequential implication of state-based reform. It now takes as much effort to answer in one state evaluative questions about program operations and outcomes as it once took for the whole country. When
the policy phenomena to be studied shift to a smaller scale (e.g., Michigan's family assistance program, not the U.S. government's family assistance program), the data needed to study any one policy shrink very little. For example, the sample size needed to describe—on a statistically valid basis—participants in a state-run program are no smaller than for a national program. Similarly, program operations information (services provided, service costs, etc.) on any one state-run program requires just as much data collection effort and analytic attention as national data once did—and possibly more if state data systems stop short of previously imposed federal reporting standards.  

The bottom line is clear: Researchers interested in tracking social policy for the nation must now go to 51 places to document policy and program operations and collect follow-up data on 51 times as many people as was previously required for a single national evaluation. Alternatively, one could maintain the historic level of research effort but now bring forth evidence for just one state—at most an eighth of the nation (if the state is California)—not the entire country. While a scale factor of 51 to 1 probably exaggerates the growth in research requirements, it conveys a rough sense of the degree to which the bar has been raised for studies seeking to examine the whole nation in a devolved world. And even to hold the growth to that scale, one must assume that local variations in programs and policies will be no greater in a state-dominated system than were state variations in a federally dominated system.

**Multiple Policy Changes**

A striking aspect of current social policy reforms is the extent to which whole sets of policies are changing at once. From ANF’s site visits over the past three years, it is clear that states are rethinking and, in some cases, redesigning social policies in a variety of areas at once,
including low-income medical and cash assistance, child welfare services, employment and training programs, child care, and child support enforcement. Multiple simultaneous changes make it almost impossible to measure the effects of individual components on the populace, because many of the programs under revision affect families—especially low-income families with children—in overlapping ways.

To assess broad-scale reforms of this sort, one must pay attention to a broad range of outcome measures and understand their interactions within an individual family. Past evaluation research on this subject generally has not tracked family and household outcomes on so many factors. With comprehensive reform now becoming the norm in the states, moving to more topically diverse data collection methods—including wide-ranging household surveys—has become essential.

Closer examination of the relationship among multiple outcomes also helps one identify the causal effects of individual policies. Simultaneous changes in three or more policy areas lead not just to more outcomes that need to be tracked but also to a vast multiplication of the "paths" along which policy might influence outcomes. Thus, changes in a given policy area may have implications beyond the household outcomes they were intended primarily to affect. Through secondary and tertiary channels, these changes may eventually cause other outcomes where other reforms have already exerted an influence. The entire complex of interconnections of this sort must be understood to isolate the effects of any individual reform component—a goal that, as discussed below, may be beyond the reach of research for comprehensive state reforms. A better strategy for examining individual reform components would be to conduct a series of random-
assignment experiments within a state that vary just one component of program reform at a time.¹¹

A more approachable analysis would look at a state’s social safety net as a whole to see if states are getting substantially different results for disadvantaged families from their different packages of social policies. As discussed in the final section of the paper, even isolating the combined effect of a state’s entire policy package can be quite difficult given the many nonpolicy factors that influence participant outcomes over time and across states. But at least then the problem is reduced to a manageable scale.

Fortunately, a lot is accomplished by studying simultaneous reforms as a package. States need to know if any progress—or regress—that low-income families experience over time stems from their reform package as a whole. If the reforms have led to progress, they should be maintained, and identifying the exact elements that produced favorable results takes second priority. However, even when disentangling the contributions of the different reform elements has first priority (e.g., where the package as a whole is found to harm low-income families, and policymakers want to eliminate the damaging provisions), one must first know the overall effect in order to have something to disentangle.

But focusing on the "whole-package" effects of comprehensive reforms has one somewhat troublesome side effect: It puts the "horse race" question—Which states are doing better than others?—at center stage. Once introduced, this question tends to monopolize the public’s attention and politicize reactions to the findings and how they are used in the policy arena, often squeezing out other important knowledge the research generates.¹² This heightened "win-lose" perception of evaluation results may be endemic to policy research in a state-
dominated world in which every state—or at least every governor—is trying to show better results than the state next door. If so, evaluators will need to work on design and dissemination strategies that minimize the "horse race" mentality and its interference with other objectives and benefits of the research.

In sum, clear thinking and correct interpretation of the role of policy in state- and family-level outcomes become much more difficult—especially for the general public—when studies must grapple with several policy reforms at once. The final section of this paper explores some of the research challenges this new situation poses, though many probably have not yet been anticipated.

**New Research Priorities**

When more basic policy questions have to be answered first, other research priorities will suffer. When studying state-based programs, the first order of business is often to determine what states are doing in the various social policy realms for which they have responsibility—how they are spending their money, whom they are serving, what policy implementation hurdles they face, and so on. As a result, still worthwhile but "farther out" questions about participant outcomes and impacts take a back seat. While these questions can still be tackled, more time and research funding are needed to reach that point compared with the era when federal agencies routinely answered questions about program operations from information received during the course of running and overseeing their programs.

An interesting offshoot of this shift in research priorities concerns the primacy of different types of data. With programmatic issues claiming a larger part of the research agenda—and findings on participants concentrating mostly on their number, characteristics, and services...
received—aggregate information on a state's spending, participants, and service delivery totals fulfills a much larger share of a study's total data needs. As a result, the individual- and family-level microdata—critical to monitoring behavioral responses to new policies and measuring their impact—enter much later in the analysis, if at all. If one wants to address these latter topics, micro-level data are still indispensable. But the bulk of the work devoted to state-by-state policy assessments may now involve state-level summary statistics. This suggests that evaluators need to put higher priority on understanding and accessing state aggregate data (from both administrative sources and national surveys) and on working with state agencies to get them to produce summary information better organized to support evaluation (in terms of time coverage, population splits, etc.).

**New Funding Mix**

Finally, the mix of funders and sponsors for evaluation research has shifted considerably in recent years, away from individual federal agencies and toward state governments, private foundations, and consortia of federal agencies. Numerous studies launched in the past few years illustrate this tendency,\(^{13}\) as private foundations,\(^{14}\) then state agencies and federal government consortia,\(^{15}\) adjusted their research agendas to focus on devolving social policies.

These are all distinctive funding arrangements compared with the prevailing model for earlier social program evaluations, in which a single federal agency provided all funding and oversight. They carry important implications for how research gets done. For example, in many cases state agencies sponsoring policy studies have considerably fewer resources to draw on than federal agencies, in terms of both funding and staff expertise and experience.\(^{16}\) Also, because they are smaller, state governments tend to have fewer buffers between political appointees and
research directors—making decisions about state research projects more vulnerable to political influence, particularly in the critical areas of identifying topics, selecting the study team, and setting time frames for the release of findings.

Private foundations and federal consortia change the equation in other ways. These sponsors have broader social goals and program responsibilities than individual federal agencies. As a result, foundations—and, increasingly, federal consortia—tend to approach issues and set their research agendas based on broad social concerns or the needs of specific at-risk populations (e.g., children living in poverty) rather than on specific government programs or reform measures. This broader focus heightens the emphasis on studying whole packages of reforms (see above) and can expand the topical breadth of the typical research project well beyond the norm of earlier single-agency evaluations. The quality of the research support and oversight that foundations and consortia provide may also differ from past practice, depending on how new this role is to individual foundations and how well multiple federal agencies coordinate their leadership.

The net result of these sponsorship changes—in terms of the content and quality of research—will not be known for some time. But already it is clear that, in relation to past norms, the potential exists for future social policy research to be:

- Undercapitalized, in terms of both finances and sponsor guidance;
- Over politicized;
- Topically sprawling and, thus, unwieldy to implement; and
- Overreaching when setting goals, compared with available research tools.
Coping with these changes presents major technical and management challenges to the research teams involved.

**ANF Features That Reflect the Changing Landscape**

In many ways, the *Assessing the New Federalism* project reflects these many changing needs and goals of social policy research in a devolved world. Table 1 shows specific examples of how the new imperatives of policy analysis—a state-by-state focus, the need to account for multiple simultaneous policy changes, a return to the basics in research priorities, and the involvement of new sponsors—have led to the particular design features of *ANF* described at the beginning of the paper. The connections between these needs and the specific *ANF* features listed are obvious by this point and so will not be detailed here.

Of course, not all of the possible implications of the new policy environment for research have been incorporated into *ANF*'s design, since *ANF* represents an evolving paradigm for social policy research that is still in its infancy. The aspects of the new paradigm that have not yet been achieved—or even conceived in some instances—will, of course, limit what *ANF* and other "first-generation” prototypes of this type of research accomplish. Design limitations in the current studies will also provide lessons for improving studies of this sort in the future. While much more extensive learning lies ahead, we can already begin to draw lessons from the *ANF* experience.
Lessons Learned to Date from the ANF Research

What has been learned so far in pursuing the ANF agenda that carries over to other similar studies? At least four lessons are evident at this early stage:

- Broad-based analyses are hard.
- Limits are essential.
- Data needs cannot be compromised.
- Clear, timely presentation of results is more important than ever.

Broad-Based Analyses Are Hard

Anything new and complex is bound to be difficult. Most practitioners would aver that the (now very familiar) research methods used to evaluate individual federal programs are difficult to accomplish, even after decades of accumulated experience. What then should we expect from new, expansively defined studies of broad social policy change? The mere scale of the subject matter tells much of the story: a range of policy areas changing at once, not just one; 50 states replacing one nation as the operative unit of interest.

The challenges of novelty and scale were well known going into ANF and other similar studies. What new indicators of difficulty have emerged from many months "in the trenches" on ANF? For starters, some aspects of the study proved easier than expected, including such goals as:

- Gaining access to key policy actors in the states when conducting site visit interviews;
- Compiling state-level variables from external sources for the state database; and
- Sustaining "cold-call" interviews through the entire 40-minute questionnaire once a household agreed to participate in the survey.
Table 1: *ANF Design Features Attributable to Shifts in the Social Policy Research "Landscape"*

**Features Attributable to . . .**

. . . The Need for State-Specific Findings

- An on-line state database that contains annual information for all 50 states and Washington D.C., compiled from external sources
- In-depth case studies in 13 states
- Household survey with large state-representative samples

. . . Multiple Simultaneous Policy Changes

- State database covering many different policy elements
- In-depth case studies covering all social policy areas and focusing on program interactions
- Cross-cutting budget analyses encompassing all social program areas
- Household survey that measures a broad range of program-related activities and outcomes

. . . A Return to the Basics in Research Priorities

- Case studies that emphasize program design and implementation in 13 states
- Tracking of welfare program parameters in all 50 states and D.C. (in state database)
- Extensive program participation analysis using state database and household survey
- Fiscal analysis that maps the flow of funds from national to state and local governments
- Limited expectations for measuring policy effects

. . . New Sponsorship Arrangements

- Broad-ranging research agenda
- Population-based, rather than program-based, sample for household survey
But challenges in other areas exceeded expectations, including such tasks as:

- Coordinating the massive staff effort involved in conducting 13 separate state case studies simultaneously;
- Assembling meaningful and comparable data in 13 states on states’ spending for children's programs;
- Overcoming initial hesitation and refusals among household interview targets, who were often suspicious of the cold-call aspect of a random-digit-dialing survey; and
- Relating disparate conceptual elements of the study to one another when developing analytic products: health topics to income and social services topics, program and policy information to data on household outcomes, lessons of past research to the potential of newly enacted policies, and so on.

A final, but not incidental, point to be made here is that old expectations of what policy studies are supposed to accomplish do not go away just because new expectations arrive. Thus, while everyone now wants to know how states integrate social programs to best assist the whole child, that information should not come at the expense of more traditional measures of the impacts of individual program components. Similarly, funding flows among different levels of government represent a defining aspect of the new federalism and must be documented, but not at the cost of solid information on what is actually done with those funds in terms of service delivery. Other examples of the "add this, but don't drop that" imperatives of state-based research could be easily added to this list, each example expanding the degree of difficulty of the enterprise as a whole.
Limits Are Essential

In light of the difficulty factors just mentioned, and particularly the relentless dynamic toward adding research components without subtracting any, the most compelling general lesson from the early stages of ANF is the importance of setting limits. For example, not every topic that could have been included in ANF was, though ideas for additions (of topic areas, research components, data sources, and states) have emerged steadily throughout the project. Given the challenges of ANF’s scope as it now stands, "holding the line" in this area has proven essential.

The same may be true for other studies. The operative rule is that scope must be contained at both the extensive and intensive margins. Strong temptations to cover additional realms of social policy will be ever present in the era of new federalism, since most state policies are changing and all interact. If ANF is any guide, researchers may also have an ever-present sense that "two more states, or three more local sites, would really put this thing over the top." Add to that the natural—and laudable—desire to do each part of a sprawling devolution study with the same level of detail and rigor as earlier, more limited studies, and the magnitude of the undertaking could become overwhelming, both in dollars and logistics.

Doing with less always has its costs, of course. This makes experience and wisdom in deciding where to draw the line as important as a resolve to live within boundaries in the first place. In the toughest situations, dropping to fewer sites or states may be the best option available when scope becomes a problem, since geographic cuts almost always make a substantial difference in feasibility and cost while compromising nothing except the share of the country to which a study speaks.17 Though the latter obviously has its cost, lowering geographic
aspirations at the outset is far better than being forced to halt or drastically shrink a study in midstream because of a lack of funds or overwhelming logistical problems.

**Data Needs Cannot Be Compromised**

Good input information is essential to any successful research project to ensure that key analyses are not compromised by unreliable or incomplete data. ANF’s decision to put data collection first among many project priorities influenced ongoing design and funding decisions throughout the first two years of the project.

It may be that the imperative to collect reliable data has even larger consequences in studies of comprehensive state reforms than in other studies. In an era of unprecedented policy freedom for states and localities and great uncertainty, basic descriptive information on policies and outcomes—the great "unknowns" of the new federalism—provides the greatest return on the dollar. In the extreme, a study that has to strip away all of its analytic and reporting functions just to bring forth new, high-quality data makes more sense than a study that skims around the edges on both data collection and analysis in order to retain both. The former makes a vital contribution by equipping others to take the next step of using the data. In contrast, synthesis and analysis may become a luxury from a funder's standpoint when something has to be cut, and to some extent this valuation is likely to be absorbed into the decision of research staff involved in the project.

This is all to the good, as long as additional resources—probably from other sources—are available on a timely basis to generate the kinds of analyses the data are designed to support. While the trade-off is seldom this severe, the balance always lies with the data. In most cases, difficult resource allocation issues are, and should be, resolved in that direction.
Over time, the trade-off between data collection needs and the ultimate goal of producing thorough, thoughtful analyses of the data may grow sharper—potentially increasing the number of times data needs impinge on other areas of an ANF-style project. Many of the types of data needed for comprehensive state policy analysis appear to be more difficult (i.e., costly) to collect today than a few years ago, while others seem likely to become more difficult to collect in the future:

- Though it hasn't been a problem yet for ANF, devolution places increasing demands on the time of key informants in state and local offices, both to implement ever more ambitious policy reforms and to respond to the barrage of researchers wanting to see them.
- Initial access and cooperation in conducting household surveys is already a growing problem, especially for the kind of population-based, random-digit-dialing surveys that may prove essential when studying the broad sweep of state policy reforms.
- The administrative data systems that states use to track program participation and service/benefits delivery are in great flux already and will change still further as they are asked to support new policy approaches (e.g., to monitor the "time-limit clock"). Over the long run, these changes should strengthen state systems as a source of research data (though some measures previously compiled at the federal level may be lost). However, in the near term, new client tracking and data reporting requirements of the Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) will make data structures and longitudinal tracking more complex.

As difficulties mount in this area, the primacy of data collection in studies like ANF may increase the pressure on researchers to make painful cuts in their research agendas or to come up with
creative designs that require fewer data to ensure that the analytic component of each state-based policy research project is not seriously eroded.

**Clear, Timely Presentation of Results Is More Important Than Ever**

The vast topic range presented by the current social policy reforms, and the prominent role of advocacy groups in the debate over those policies, ensure that information and recommendations on the subject will inundate decisionmakers for the foreseeable future. As advocates and political organizations turn more and more to research (often their own) to make their cases, less partisan findings from independent researchers may have a more difficult time getting noticed. Even within the set of rigorous nonpartisan studies, the market will be quite crowded and the flow of new information to consumers fairly overwhelming.

In this clattering world, *how* information is circulated—and its accessibility to the vast and growing state and local audience—will determine the influence of individual research projects as much as a study's content or institutional affiliation. As a result, researchers can no longer assume that doing the right work, then releasing it through standard channels, will place their findings in front of the key decisionmakers. Two things have changed: thanks largely to *USA Today* and the Internet, packaging and media matter more than ever; and audiences have grown and diversified. Sending long, scholarly reports to top national opinion makers, federal officials, and research colleagues now constitutes just a first step—and perhaps not a very good one. As the nexus of decisionmaking moves from federal officials to states and local communities, dissemination must follow, using new types of products and expanded mailing lists.
The timing of research results has also grown more important, heightened by changes in the policymaking nexus. States, more than federal agencies, like to refine social policies and programs continuously, always trying to move to a more effective version. This updating creates a constant demand for timely information to guide policy, with a strong preference for highly up-to-date information that tells policymakers exactly where they are now while they are still picking the next thing to change. Though no research agenda or dissemination strategy could possibly meet all such desires, the standard practice in large-scale evaluations of asking policymakers to wait years for usable findings will become increasingly untenable when doing research for a largely state-based audience.

The Final Challenge: Linking Policy to Outcomes

One challenge that still lies ahead for ANF and other comprehensive assessments of new state-based social policies merits in-depth attention: the problem of attributing social outcomes to the specific policy choices made by states. To attribute outcomes to policies, two essential evaluative questions must be addressed:

• When outcomes improve, how much of the improvement comes as a result of the policies put in place by state and local governments, and how much would have occurred anyway (i.e., even absent policy reform)?

• Of the new policies and programs attempted, which ones work the best—that is, make the greatest positive contribution to social outcomes in their states?

The first of these questions concerns the net impact of the new programs, compared with the old; the second concerns the "differential impact" of various new programs. Only by addressing these
two topics can a study achieve the ultimate goal of policy research: to discover which policies and programs contribute most to advancing society's goals.

To provide this information, a study must somehow distinguish between policies that simply accompany better outcomes without causing them and policies that make a causal contribution to those outcomes. If all other things except the policy component can be set equal when outcomes are compared (either between states or over time within a particular state), any differences in results are appropriately attributable to policy differences.

The Challenge

Anyone who has ever tried to attribute outcomes to policy knows how hard it is to equalize nonpolicy differences in cross-state or cross-time comparisons. Social experiments achieve this "all other things equal" goal—and thus provide highly reliable measures of program impact—by holding time interval and location constant but randomly assigning individuals to different policies. When large numbers of people are picked at random to receive either policy A or policy B, the two groups will not differ on average except through the effects of the policy.

Researchers interested in the effects of comprehensive state reforms do not have the option of equalizing other factors using random assignment. Instead, they must take different state policies where they find them—policy environment A in state X, policy environment B in state Y—and wrestle analytically with the fact that the two states differ on a number of other important factors that influence outcomes. This pushes the analysis into the realm of nonexperimental methods for attributing population outcomes to policy choices. The remainder of this section explores some of the more promising possibilities in this realm, though it is
already clear that no easy solution will be found. Indeed, there is some chance that all existing methods will produce results too implausible or ambiguous to credit.\textsuperscript{19}

Notwithstanding the qualifications that may apply, a number of techniques for attributing outcomes to policy following comprehensive reform are worth considering and debating at this point. All of these techniques flow from a common framework for viewing the world in the reform era. At the abstract level, this framework consists of (1) a set of (perhaps heroic) assumptions about the data that will be available to do the analysis and (2) the theoretical construct of a counterfactual, or alternative, world in which the policy reforms of interest have \textit{not} taken place. The framework can be stated as follows:

- Outcomes for low-income families and individuals will be observable at the national, state, and sometimes local (county, metropolitan) level,\textsuperscript{20} either in the aggregate for a given geographic area or for a sample of individuals who statistically represent the area.

- The social policies and programs operating in those areas will also be observable, including state-by-state variations in policy, as well as county-by-county variations in some instances.

- Both outcomes and policies will be measured over several years, beginning before implementation of the reforms and continuing several years beyond.

- To measure the contribution of policy reforms to outcomes, one must posit and then quantify what the outcomes would have been without the reforms. This is the counterfactual, against which the accomplishments of real policy reforms are measured.

- To measure policy influences, outcomes in the counterfactual world must be estimated using observed outcomes for people, localities, or time periods where reform had not taken place. These outcomes are then adjusted to make them more similar on nonpolicy factors to
observed outcomes for the people, time, and state where the reforms did apply; then the two sets of outcomes—the "with reform" and "without reform" scenarios—are compared. Any differences are interpreted as the influence of the policy reforms alone, since other factors have been equalized.

Possible Responses

Within this framework, a number of strategies either have been suggested previously for measuring the contribution of policy reforms to outcomes or can be advanced here for the first time. Most are well-known techniques for measuring the effects of broad-based policy reforms that influence both individuals and the larger social system. These techniques have been used many times in the literature on program evaluation; identifying their originator(s) or citing their most important applications is beyond the scope of this paper. The focus here is narrower, asking "What techniques have potential for accurately attributing outcomes to policy in the context of comprehensive reform of state social programs?"

At least four basic strategies seem worth considering at this point:

- **Pre-post, cross-sectional comparisons** across a range of states, each tracked over a series of years during which policy reforms are implemented, with prereform years or nonreform states serving as the counterfactual;

- **Comparison group analyses** that contrast the low-income families most likely to have been affected by social program reform to similar individuals not affected by those changes;

- **National** comparison of observed postreform outcomes to a simulation-based counterfactual of the nonreform world; and
• "Theories of change" analysis that verifies intermediate steps in the postulated causal chain that links a policy change to population outcomes.

Pre-post, cross-sectional comparisons may be the foundation for any in-depth examination of the policy-outcome connection. This approach looks simultaneously at the two key evaluative measures mentioned above: net impacts compared with prior policies in a state, and differential impacts among alternative reforms in different states.

All recent analyses of welfare reform's contribution to the Aid to Families with Dependent Children (AFDC) caseload decline use this approach, which essentially treats each year in each state as a separate observation and explains outcomes as the joint result of the policies in effect in that state and year and other state- and year-specific factors such as long-term historic wage levels in a state or national social norms during the year. The policy variables in the model—which can vary both across states and over time within a state—are either 0/1 indicators of when and where different pre- and postreform policies (e.g., AFDC work sanctions) were in place or continuous measures of the extent or "strength" of a given policy provision once implemented (e.g., the share of AFDC recipients sanctioned during a year).

The basic pre-post, cross-sectional approach can be enhanced in a number of ways. Fein (1996) leaves the policy variables out of the model when choosing the contextual variables needed to control for state- and time-varying factors other than policy. This allows him to maximize the portion of his outcome measure (AFDC caseload size) accounted for by contextual factors without running the risk of selecting a final model—or appearing to select a final model—based on the nature of the policy results obtained (i.e., "fishing" for the most desirable results). Others, such as the Council of Economic Advisers (1997) and Martini and Weant
(1997), allow policy measures from year T to influence outcomes in years T – 1 and T + 1—the former because the announcement of (or even just the threat of possible) policy changes could change behavior, and the latter because some policy reforms may take time to have an influence. Validation of the model for the pre-PRWORA period, comparing its results to those produced by random-assignment evaluations of 1115 waiver reforms, could also prove worthwhile. The goal here would be to identify adjustments to the pre-post, cross-sectional framework that move its findings closer to the rigorous, highly reliable results provided by the waiver experiments. If this is achieved in the pre-PRWORA interval, we would expect the adjusted approach to be more reliable for the post-PRWORA period as well.

**Comparison group analyses** could potentially be grafted onto the basic pre-post, cross-state analysis. The largest concern in relying on pre-post, cross-state models is the possibility that the relevant outcome in a state could take a turn (up or down) unrelated to policy reform but in the same year that reforms are implemented, while other states do not experience the nonpolicy shift. A pre-post, cross-sectional design would have no hope of detecting a state-specific shift of this kind, or of separating it from the influence of a simultaneous change in policy, unless the underlying shift concerns factors that are themselves measured and controlled for in the analysis (e.g., changes in a state unemployment rate). Unfortunately, many factors not included in the model could put a state on a new course at an inopportune moment and confound pre-post, cross-state indicators of policy effects.

To avoid this problem, one needs a comparison group in the same state, followed over the same time interval, to reveal any shifts in underlying trends. Where can such a comparison group be found, specifically one not affected by policy changes? When studying the AFDC reforms,
one option is to look at single-parent families within the state that are not on welfare. At any time, there are a large number of families in this population, including former welfare recipients. However, there is a considerable possibility that welfare reform will change long-term options and behavior for these very families, which may think they will need public assistance at some future point. Expectations of confronting a different public assistance program in the future could change the immediate actions and financial situation of these families, even though not currently on welfare, contaminating them as comparison group members. More "insulated" comparison groups, though ones less likely to mimic the welfare population in other respects, might also be found if this notion is pursued further. Statistical techniques for incorporating comparison group data into pre-post, cross-state impact models also warrant further attention.

The idea of a simulation-based counterfactual may be new to the literature. The basic intuition here is that any simulation system capable of predicting future outcomes under different policy scenarios can surely do a decent job of predicting future outcomes under the policy scenario in which it was derived (i.e., the actual policies in place during the period when the model was estimated). Thus, the 1996 version of such a system, such as the Urban Institute's TRIM2 system, should give a good indication of what 1997 and 1998 outcomes would have been had policies not changed during that time, particularly once 1997 and 1998 contextual data on labor market conditions and other nonpolicy factors are plugged into the model. While this may not help in analyzing policy impacts for individual states (e.g., there is no state-specific version of TRIM2) and cannot be used to compare one state's new policies to another's, it does offer
possibilities worth exploring when analyzing the net impact of the nation's devolution reform movement as a whole.

A final analytic approach that seems to hold promise in understanding the causal effects of state policy reforms is the "theories of change" technique, which verifies intermediate steps in the postulated chain connecting policy to outcomes. A substantial literature on this method, and a number of past and ongoing applications, attest to its appeal. Basically, the "theories of change" approach looks to see if the precursors to policy impacts are visible in the data where theory says they should appear. Suppose, for example, that one finds less juvenile crime in states with higher spending on school-to-work employment programs. Before concluding that the additional spending caused the fall in crime, one had better confirm that (1) more spending in fact means more youths served by school-to-work and (2) more youths served translates into more youths working (the one means by which employment assistance is expected to translate into less crime). Where these linkages cannot be confirmed, observed correlations between policy and outcomes should not be interpreted as causal effects.

**Next Steps**

The strengths and weaknesses of these various analytic methods need to be explored at much greater depth than is possible here. So, too, should the potential for combining the best features of each strategy in a multilayered approach to sorting out policy-outcome linkages. Other strategies for linking policies to outcomes may be added to the list—and some of them implemented—in the work on policy-outcome linkages that lies ahead for the *Assessing the New Federalism* project and other devolution studies that address the same issues. Finding a way to accurately and convincingly attribute outcomes to policies is a goal deserving of the highest level
of attention—and collaboration—among policy researchers and evaluators for the next several
years, as the results of social program devolution unfold and everyone wants to know what got us
there.
Notes

1. State decisions also provide the context for local reforms and occur sooner than policy changes in local communities.

2. Examples include the National Job Corps Evaluation, the National Job Training Partnership Act (JTPA) Study, and the JOBS Evaluation, among many. To the author’s knowledge, a complete list of all such studies does not exist. Greenberg and Shroder (1997) provide a comprehensive summary of employment and training program evaluations using random assignment to measure program impacts (i.e., social experiments). See LaLonde (1995), U.S. Department of Labor (1995), and Friedlander, Greenberg, and Robins (1998) for recent summaries of the findings from this literature.

3. For example, the history of program evaluations presented here does not necessarily apply to early studies of income transfer and health policy innovations, such as the Income Maintenance Experiments and the Health Insurance Experiment. Nor do the conclusions here apply to all of the more recent employment program evaluations—or to all the government agencies that sponsor them.

4. For examples, see Greenberg and Shroder (1997).

5. For example, the nation's largest job training system for disadvantaged workers—programs authorized by the Job Training Partnership Act—collects 13-week follow-up data for individuals leaving the program.

6. The remainder of this section assumes that a good deal of devolution occurs in practice in crucial social policy realms, not just in principal (i.e., in legislation). Measurement of the actual extent of federal-to-state devolution in various dimensions (control over funding, autonomy in program design, etc.) is the first step in any study of the new federalism.

7. It is important to note that some of the changes listed here began years ago for selected policy evaluations, such as the Manpower Demonstration Research Corporation's (MDRC’s) evaluations of state-based welfare-to-work demonstrations in the 1980s and the Section 1115 waiver evaluations of state welfare reform initiatives in the 1990s. What is different today is that all, or nearly all, social policy evaluations will have to grapple with these changes, as more and more research topics of primarily federal concern devolve to the states.

8. Section 411 of the 1996 welfare reform bill (the Personal Responsibility and Work Opportunity Reconciliation Act, or PRWORA) requires extensive state data collection and reporting by states on both overall program characteristics and individual participant characteristics for the new federal family assistance block grants, Temporary Assistance for Needy Families (TANF), following uniform national standards. It remains to be seen how successful individual states will be in implementing these standards.
9. There are some fixed costs involved, regardless of the number of states involved or the overall scale.

10. Many recently initiated studies of welfare reform take the alternative view—that devolved social policies cannot be understood sufficiently without looking primarily at the local level. Several of the most visible multisite studies of welfare reform, such as MDRC’s Urban Change project and the Multi-City Study of the Effects of Welfare Reform on Children led by Johns Hopkins University, focus on individual cities. Several factors may explain this emphasis: first, a sense that, in welfare reform, the toughest challenges and the majority of participants come from urban areas; second, the possibility that some important elements of devolution may bypass state government and move straight to the local level; and third, a desire to understand the fates of low-income families in the context of their local communities. In contrast, ANF focuses on states as the most relevant policy unit for tracking the first step of devolution (federal to state), though the local orientation has not been completely omitted (witness the extension of the case study research to two or three local sites, both urban and nonurban, and the inclusion of a large local survey sample in Milwaukee).

11. This strategy is advocated by many observers, most notably Robinson Hollister.

12. Continuing the “horse race” example, findings that allow the public to focus on which horse is leading the race and which is lagging—and to speculate on which jockeys or trainers might get fired as a result—can get in the way of more fundamental lessons about horse racing drawn from the same study (e.g., what kinds of horses run best, which training regimens produce the best race-day results, and how track conditions affect performance).

13. The list of new studies is quite extensive. Descriptions of most of these studies appear on the Internet courtesy of the Research Forum on Children, Families, and the New Federalism (http://www.researchforum.org). A more basic list is available from the author.

14. Besides Assessing the New Federalism, other studies of large-scale social programs initiated recently with foundation funding include the Center for Health Systems Change at Harvard, the Multi-City Study of the Effects of Welfare Reform on Children led by Johns Hopkins University, and the Project on Devolution and Urban Change at the Manpower Demonstration Research Corporation. Many other similar foundation initiatives have begun.

15. Examples of state-initiated welfare reform research include the "Track 2" evaluations of new state approaches authorized by PRWORA, ongoing (and often expanding) 1115 waiver studies from the pre-PRWORA era ("Track 1"), and various state "data archiving" initiatives such as the UC-DATA project at the University of California. Consortia of federal agencies also have formed to track child well-being indicators (the Federal Interagency Forum on Child and Family Statistics), evaluate the effects of new welfare-to-work funding in the Balanced Budget Act of 1997, and complete other similar projects.
16. One of the strengths of the federal agency approach to sponsoring social policy research is the long-term institutional and staff learning that goes on in offices dedicated to this purpose. Many government project officers specialize in directing research studies of a certain kind and know a great deal about which research objectives are attainable given available methodological tools. As a result, these offices have a history of using well-crafted analytic strategies to obtain highly conclusive results on only the most "answerable" policy questions—and usually just one or two (or at most a handful of) policy questions per study.

17. Unlike other types of cuts, what's not done in a limited-site study cannot possibly eviscerate what is done, since no essential components have been omitted for the sites covered. In contrast, cuts in topical coverage or data collection within a site (or for a whole project) can make all results incomplete and unsatisfying.

18. Random assignment could be used to evaluate policy variations within a state, if a state (or even a set of counties) is willing to run parallel programs in the same locale and assign individuals to the different programs at random. This was the dominant technique used in evaluating state welfare reform initiatives prior to PRWORA (those initiated under 1115 waiver authority), and many of those studies continue with their randomized "treatment" and "control" groups intact, though often on a fairly small scale.

19. See, for example, a recent spate of papers estimating the contribution of states' 1115 waiver policy reforms to the sharp reduction in AFDC caseloads experienced during the mid-1990s (Fein 1996; Council of Economic Advisers 1997; Ziliak et al. 1997; Blank 1997; and Martini and Weant 1997). Some of the limitations in technique and ambiguity in results on this topic are noted by Martini and Wiseman (1997) in their examination of the Council of Economic Advisers report.

20. In an analysis that focuses on state policy changes, the availability of state-level data is absolutely essential. Information for the nation as a whole or for localities within states may not be needed in certain cases.

21. An oft-suggested approach is to use former welfare recipients at the time they were on AFDC as the comparison group for examining current, postreform participants. For example, the MDRC Urban Change project will feature "cohort" comparisons of this sort. Given the overlap between the past and current welfare recipient populations—especially within the crucial set of long-term recipients—and the need to remove differences due to nonpolicy time trends, this strategy presents a number of analytic challenges. Even so, it carries enough promise to make it worth investigating.

22. See Scheirer (1994) for an overview and Chen (1990) for a more complete treatment of the subject.
References


About the Author

Stephen H. Bell is a principal research associate in the Income and Benefits Policy Center of the Urban Institute. From 1996 to 1997, he served as director of research for the Assessing the New Federalism project, and he continues to pursue research in this area. An economist by training, Dr. Bell specializes in quantitative methods for estimating the effects of employment and training programs and welfare reform on participants. His work on evaluation methodology includes such publications as Improving the Evaluation of ETA Pilots and Demonstration Projects, "Are Nonexperimental Estimates Close Enough for Policy Purposes?" and Program Applicants as a Comparison Group in Evaluating Training Programs.