

# **Research and Evaluation Methods in the Field of Social Welfare: Weighing in on Their Use to Address Poverty**

**By Rebecca May  
1999**



**Center for Fathers, Families, and Public Policy  
23 North Pinckney Street, Suite 210  
Madison, WI 53703  
608-257-3148 (Tel)  
608-257-4686 (Fax)**

# Research and Evaluation Methods in the Field of Social Welfare: Weighing in on Their Use to Address Poverty

By Rebecca May

## Introduction

An impact analysis finds that its research subjects were significantly more likely to earn slightly higher incomes one year after implementation of a welfare reform effort that required job search activities leading to employment. In a distant county of the same state, a concurrent program that encouraged G.E.D. attainment prior to job search yields less caseload decline and lower income in the same time period. Interest in moving welfare recipients quickly from welfare to work increases as states initiate their own welfare reforms, often guided in these efforts by such research results. The President endorses work-first, initiating dramatic changes to welfare, including time-limited assistance and state flexibility to restructure welfare programs in almost limitless ways. Caseloads decline nationally, but a sizable number of former recipients are missing. They don't appear in welfare offices, aren't increasing the demand for childcare, and certainly don't appear to be moving out of poverty. Food banks report increased demand and homeless shelters overflow, but the reports are anecdotal and not easily traced to welfare reform.

The connection of research to poverty and hopelessness is not often at the forefront of welfare policy debates, but it is to this connection that this edition of *Issues and Insights* hopes to speak.

As one subject of the Center on Fathers, Families, and Public Policy (CFFPP)'s ongoing colloquia and *Issues and Insights* series, we focused on social welfare research methods that are predominant in the evaluation of policies and programs targeted at the poor. We began with a colloquium discussion, some of which will be incorporated here. One goal of this issue is to encourage practitioners, whose clients are the intended beneficiaries of research, as well as those who make policy decisions regarding people in poverty, to take a more critical look at some of the assumptions inherent in many of the most accepted research methodologies.

The research community is looked to by policymakers, legislators, practitioners and the general public as an unbiased source on which to base opinion and policy. As the reliance on large-scale impact evaluations has grown, so have the funding and reputation of the research industry, with increased acceptance that its methodologies are preferred as a "pure" source from which to gain irrefutable information prior to applying that information to new political initiatives. But this search for straightforward answers through research experiments and data analyses, although it has yielded information on isolated factors that contribute to poverty, has resulted in little policy or program innovation that can be credited

with changing the economic prospects of the poor in a meaningful way. Indeed, one could argue that our understanding of even the most basic questions, such as who is actually poor and why, has not been clearly advanced through many of the current poverty research methods.

The relationship between poverty research and public policy creation has been an intimate one since the mid-1960s, when poverty emerged as a central social issue during President Johnson's "War on Poverty," and research organizations were created with the express purpose of informing policy-makers (Wolfe and Corbett, p. 2). Once its efforts to aid the poor were underway, the Johnson administration determined that its new social programs should be studied and evaluated to determine their effectiveness. New government programs were considered difficult to justify without "firm data and systematic thinking," and this fueled the 1965 executive order directing all federal agencies to incorporate measures of cost effectiveness and program evaluation into their decisions (ibid.).

In the ensuing debate over what would constitute sound evaluation, non-scientific methods were all but abandoned in favor of economic models, experimental designs, and survey instruments. These models held the promise of providing "value-free" tools along with the appeal of objectivity in a politically charged climate. The New Jersey Income Maintenance Experiment, conducted in 1968 and designed to study the behavioral response to varying minimum income guarantees, was perhaps the first project to use random assignment. The New Jersey experiment helped to "make the methods of experimentation formerly reserved for natural scientists available to social scientists" (Wolfe and Corbett, p. 2).

Thus a mutually beneficial system was created: government policy toward the poor was legitimated by the research industry it helped to create, and the research industry was afforded a reliable source of prestige and funding by the government for investigating the behavior of poor families. Experimental designs employing random assignment of program participants evolved to become the standard method used to evaluate the effectiveness of social welfare policies and programs. The opportunity for a broad range of research methods to be equally accepted and funded might have been lost as a result. Manski and Garfinkel have noted that "present contractual funding encourages assembly-line evaluations, executed with conventional procedures, reported in a standardized format. It discourages innovation in methods, efforts to understand the complex set of processes that define a program, evaluation of long-term program impacts, and creative thinking about program design"(Manski and Garfinkel, 1992).

### **The Experimental Design**

Most evaluations of state welfare reform demonstrations employ an experimental design, the preferred method of research by the Department of Health and

Human Services (Gordon, 1996). With the Family Support Act of 1988, Congress not only mandated the implementation and evaluation of the Job Opportunities and Basic Skills training program (JOBS), it stipulated that "a demonstration shall use experimental and control groups that are composed of a random sample of participants in the program" (Manski and Garfinkel, 1992). Federal rules on state "waivers" for AFDC programs also required experimental evaluation (Schram, 1995).

The core of the experimental design is random assignment.<sup>1</sup> Random assignment is considered a necessary tool for the analysis of welfare reform because it allows for the establishment of a comparison group that is statistically identical to the clients receiving services in every way except for the actual receipt of those services. Any other method for establishing the comparison group becomes quickly bogged down by the difficulty of interpreting the results, since other differences between the comparison group and the participants could explain program impacts. The goal of random assignment is to isolate variables and speak definitively to the impact of a policy or program on the isolated variable.

Driving the reliance of research on random assignment is the attempt to establish that both taxpayer savings and welfare recipient self-sufficiency have been achieved as a result of a particular reform effort. But the widespread acceptance of this research approach, and the use of its results to inform policy, overlooks some important shortcomings:

- Regardless of researchers' warnings that their results speak only to the particular program model and environment studied, in practice results are often generalized, and variables found to produce significant impacts in one study are applied to other sites and program designs.
- When programs adapt themselves to accommodate an experimental design, the changes alone may diminish the programs' effectiveness. Recruitment, selection, and services offered must be standardized and consistent, but effective programs may need more flexibility and autonomy than is allowed under these requirements (Schorr, 1995).<sup>2</sup>
- There is often little interaction with the subjects by the researchers conducting large-scale quantitative research. This lack of interaction can result in research that ultimately leads reform efforts in directions that are inconsistent with the needs of low-income families.
- Perhaps most important, an experimental design may produce reliable information regarding impacts but cannot be as rigorous in explaining why the impacts occurred. The "spin" that is put on the results is no less subjective than in other forms of analysis, but it is assumed to be objective because it is derived from objective data. This is particularly troublesome given that the research firm in most cases either has been funded by the governmental entity that is being evaluated, or has created the demonstration itself and thus has a vested interest in showing positive

results. After having been entrusted with funds and faith that the evaluation will provide an important contribution, evaluation firms are understandably under pressure to find something significant. Executive summaries of such reports rarely lead with a pronouncement that nothing much seemed to have been affected by the evaluated program.

Within the experimental design itself, there are several important points at which the rigor of the study is compromised:

- The ability to generalize beyond the sample is limited by the site selection, which is rarely a random process.
- People, caseworkers in particular, respond to non-evaluated programs in a different way than they do to programs with an evaluation in place, since they cannot be kept unaware of the research status (control or experimental) of subjects as would be the case in a medical model (Manski and Garfinkel, 1992). Random assignment prevents caseworkers from choosing members of the experimental group, but it cannot ensure that the treatment received by the experimental group is not enhanced or targeted to individuals most likely to be successful. The participation of a site in a large-scale evaluation gives program staff, from managers to caseworkers, a vested interest in looking effective compared to other sites.
- The control group may not represent a true comparison because the "counterfactual" cannot be established. Welfare reform may have changed the program under study too dramatically to allow for a control group that represents the program prior to reform.
- Unintended effects of a program, and in some cases the conflict of goals among programs, cannot be evaluated using the experimental design because the variables have been isolated to a point that their usefulness is limited (Mayer, 1996). For example, attempts by Head Start to increase parental involvement conflict with work requirements under welfare reform, but the impact of one on the other cannot be ascertained through an experimental design.

### **The Control Group**

One of the primary concerns with the predominant use of the experimental design is with the control group. The existence of a control group requires that every quantitative study wrestle with the ethical issues inherent in denying services to a random group of economically vulnerable people. Researchers often justify random assignment with the assertion that the control group is no worse off for not receiving a service or benefit, and is sometimes better off if the treatment is a penalty. But, because members of the control group are often made aware of the program prior to random assignment, they are by definition made aware that there is something available to others but not to them, simply because of research that has little value to them.

One ethical argument for random assignment is that under most circumstances it is conducted only on "mandatory" participants under threat of losing benefits. The expectation is that recipients who are required to participate in a reform will not be harmed by an assignment to the control group since, as a member of a control group, they will be subject to fewer program demands. But this distinction does not account for persons motivated to participate even though the program is mandatory. Those who have volunteered to take part in a program (even when participation is mandatory), and who then find that they are assigned to a control group are frozen out of a service they wanted and needed. For certain individuals of this already vulnerable group, such exclusion could have significant repercussions and could become the last time they try to participate in a program or trust the sincerity of any program offering services to them.

The assignment of participants to control groups has at times crossed an even more questionable ethical line. In an effort to compare components of a national employment and training program (the Job Training Partnership Act, or JTPA) the Manpower Demonstration Research Corporation (MDRC) created control groups only after participants volunteered, went through an eligibility determination process *and* completed an assessment process that sometimes lasted several weeks. Once the appropriate category of services for the participants was determined, random assignment took place, excluding controls from services through JTPA for a period of eighteen months (Doolittle and Traeger, 1990). The difficulty of carrying out random assignment on persons who had expended this time and effort led many JTPA sites to opt out of the study and made site recruitment a particularly challenging, and not random, process (Hotz, 1992).

In another instance, Parents' Fair Share (PFS), a large-scale demonstration program with an evaluation component, drew its sample from noncustodial fathers who were delinquent in their child support payments but who had not yet been pursued by the child support administration. The practice of not pursuing these fathers was often an informal policy; when offices lacked the necessary resources to track the men down they expected little return in the form of child support payments. The demonstration itself, however, required that these files be "worked," meaning that additional effort was given to finding the nonpayers. For those who were brought in and found eligible for program services by virtue of their lack of employment, random assignment was conducted (Doolittle and Lynn, 1998). At this point, noncustodial parents were either assigned to the PFS program or assigned to the control group. The control group was not eligible for PFS services that included assistance in finding employment, peer support groups, and a reduced monthly child support obligation during program participation. This process created two problems not acknowledged in the PFS reports, one of which is ethical and the other methodological.

First, because the controls might have been "left alone" and not pursued by child support if not for the study, the study itself changed a practice (overlooking

certain cases rather than enforcing them) that had been an informal method for contending with impoverished fathers. Fathers assigned to the control group, by virtue of their having now been located, were subject to harsh penalties for nonpayment, which would not have been inevitable except for the study, and also were excluded from PFS program services.

Second, this same process calls into question the contention that the control group represents a true comparison, since its members were not treated in the same way (i.e., not pursued by the child support office) as they would have been in the absence of the study. While it is possible that some of these fathers may have been pursued eventually even without the study in place, it is not possible to know whether or not the practice of neglecting the cases would have been maintained.

Given the negative impact that being assigned to a control group can have on an unknown number of people in the group, it seems incumbent on researchers to explicitly weigh the value of the prospective research results against the possible cost to this group. Because the control group has become an unquestioned instrument in evaluating policy, this issue needs to be more explicitly addressed.

### **Current Welfare Reform and Random Assignment**

One of the most significant changes brought about by the Personal Responsibility and Work Opportunities Reconciliation Act (PRWORA) is an increased emphasis on individual behavioral requirements as part of program eligibility (Blank, 1996), which will in turn change the terrain for researchers in the near future. Eligibility is now contingent on the individual's conforming to a state's prescription for personal behavior. States have implemented such requirements as time-limited welfare, limits on grant amounts for additional births, cooperation with child support enforcement, and the introduction of personal responsibility contracts. Each of these policies represents an attempt to manipulate personal behavior toward a value-laden ideal of married families with a working head of household.

The resulting decentralized and subjective decision-making on the part of welfare caseworkers poses particular challenges for researchers. In programs with such complex variables and goals, isolating impacts is yet more difficult and more costly. Research may become increasingly narrow, attempting to isolate the smallest of program impacts on human decisions and behaviors.

Another result of recent reforms is that, because they are so dramatic and specific to a particular state, they may even render the discussion on the effectiveness of controlled experiments temporarily moot. Time limits, work requirements, and the array of new rules for all welfare recipients mean that there is no control group that can be safely said to have been unaffected by the reforms typically evaluated using an experimental design. Additionally, the variety

of components makes it nearly impossible to separate out impacts and be comfortable that the interplay of component goals has not affected outcomes.

The temporary puzzle produced by the reforms has by no means led to a paradigm shift in research methods, however. The experimental design remains predominant in the field, with "rigorous research" nearly synonymous with random assignment. Many experimental designs are just beginning implementation; as the reforms stabilize and remain in place long enough to form a comparison, the research industry will no doubt resume widespread implementation of experimental designs.

## **Ethical Review**

Participants in our colloquium confirmed that there is no established federal ethical review process for research on human subjects when the subjects are welfare recipients. In academic settings the review process is stringent; little research can be conducted without meeting ethical standards. One participant noted that his state and a major university attempted a collaboration to study welfare clients, but the university expressed ethical concerns regarding random assignment. When the Department of Health and Human Services was consulted to determine what ethical guidelines it had in place, it was learned that no established ethical protocol for welfare recipients exists: the consent to participate in research is implicit in the receipt of public assistance. In addition, national foundation representatives reported that there is no established standard for ethical review of research methodologies on the part of grantees, but that the grant-making process itself considers ethical issues along with all other criteria in funding decisions.

The privacy of low-income people is too often deemed by researchers to be of little value. Many welfare studies require subjects to divulge extremely personal information and to share details of their lives considered private in most other venues. In discussing the freedom of researchers to experiment on the poor without the constraint of ethical guidelines, Stanford Schram (1995) notes:

There is an insidious, if unintended, subtext to the rise of this new "social experimentation." Welfare policy research has been implicated in a pernicious but pervasive logic: the poverty of poor people is a mysterious thing, attributable in good part to their individual behavior, worthy of being medicalized in terms of schemes for changing their behavior. Policy makers might have a hard time justifying this sort of nonconsensual "experimentation" when providing benefits for the nonpoor; however, there is little political opposition to treating the poor as a special group for whom such experimentation is appropriate.

## The Use of Research to Inform Policy

In large part it is the summary of research findings, and not the details found within the report, that informs policymaking and analysis. This helps to explain the appeal of quantitative research. It is much easier to summarize quantitative results in a concise way that is easy to remember and disseminate than is the case for more qualitative studies.

An evaluation based on an experimental design can provide only a snapshot of an evolving and complex story, but its summarized results can have a significant impact, through policy recommendations and subsequent reforms, on a poor family's options. Although the experimental group may eventually earn more than the control group, the increase in earnings may tell only a small part of the story. Many impacts are not quantifiable, may take longer to unfold, or may be too subtle to detect at a given point in time. In fact, the current "work first" principle of welfare reform is based, in part, on the evaluation of GAIN<sup>3</sup> -a program implemented in Riverside County, California, that made job search the priority activity for participants, regardless of individual educational needs. This strategy resulted in more earnings for recipients in the short term compared to Tulare County, where welfare recipients were geared to educational activity (Riccio, et al, 1994). The inclination to look for a "winner" and "loser" in an impact study of this kind led policymakers to seize on the job search model as the most effective means to reduce welfare costs and increase self-sufficiency, despite a lack of information about long-term and non-quantifiable impacts.

Even the most rigorous of large-scale studies must struggle with what to recommend to policymakers, particularly when the results show only a minimal impact or no impact at all. For all the painstaking effort that goes into "purifying" the research design, the interpretation of the results is still subjective. In demonstration programs in particular, there is a vested interest in the outcome of the research. It is noteworthy that at CFFPP's colloquium on this subject, there were several impact studies described by persons involved in both the programs and their evaluation, and for each one there was no significant difference found between the experimental and control groups on the study's most critical variables. Yet in each case, the researchers were confident that the program did have an impact on its clients, but that the study failed to detect it. The lack of demonstrated impact was attributed to poor data, small sample sizes, problematic research methodology, or an incomplete evaluation. However, in every case, researchers did not waver in their conviction that the program worked despite the results.

In spite of the often equivocal results of impact evaluations, studies that do not follow the predominant model have a less clear impact on policy formulation. Katherine Edin and Laura Lein's work, *Making Ends Meet: How Single Mothers Survive Welfare and Low-Wage Work*, may represent an example of the difficulty of persuading policymakers through alternative forms of research. In their work,

379 single mothers were interviewed about their finances and barriers to surviving on either welfare or low-wage work. The findings could provide useful and important information to states as they implement strategies to move this population off of welfare and into low-skilled jobs, but the method of sampling, which was not random, has limited its influence on policy.

There has been a prevailing assumption in the field that only through an experimental design can bias in research be avoided. In discussing a comparison of nonexperimental and experimental findings, Gordon, et al (1996), of Mathematica Policy Research, Inc., state that the most common explanation of differences in the findings of these two forms of research is that the "nonexperimental estimators" are biased because they provide different results than would be found in an experimental evaluation. Thus, they conclude, experimental estimators are preferred. "DHHS shares this conclusion," they state, "as shown by the strong preference it exhibited for experimental evaluations of the welfare reform waiver demonstrations."

This rationale for relying on quantitative methods, that they are less biased and are preferred by the government, seems to be the prevailing wisdom in social welfare research. Manski and Garfinkel point out the following:

The desire to increase the general credibility of program evaluations has been largely responsible for the recent decline of interest in structural evaluations based on observational data and the associated rise of interest in reduced-form experimental evaluation. The evaluation literature using observational data has, for the most part, been explicit about its structural assumptions. In this literature, it is common to find soul-searching and debate on the validity of assumptions and on the sensitivity of results to variation in assumptions. The literature using experimental data, on the other hand, has tended to leave its assumptions implicit and unquestioned. (1992, p. 19)

## **Conclusions**

This paper's focus on research derives from a perception that it is not possible to create adequately informed welfare policy without looking carefully at the research methods that are used to inform us in the first place. The narrow focus of social welfare research on changing the behavior of the poor in order to reduce poverty has played a little-debated but critical role in defining the locus of reforms. Current welfare reform efforts are a perfect case in point. Absent from the debate in the creation of time limits, family caps, and work requirements was a more than cursory consideration of the availability of quality employment and the willingness of private sector employers to hire welfare recipients.

The track record of the social welfare research industry does not merit the current reliance on the experimental design, or on the status quo of research, as the appropriate place to invest hopes of finding solutions to persistent poverty.

Considerable debate persists about the most fundamental questions that the research industry has attempted to address over many years and after untold expense. Reliance on the experimental design may result in glimpses of hope here and there for a narrowly defined intervention, but the interventions and their impact must first be boiled down to an isolated piece of the complicated poverty puzzle. One cannot much hope for a sea change in conditions, even when research has led to such a change in policy, based on such findings. The policies that are created in this research environment tinker with program requirements while attempting to finely hone the behaviors of the poor. They are not generally guided by a belief that the systemic barriers to overcoming poverty must be addressed, but focus instead on the perceived shortcomings of the poor themselves.

---

<sup>1</sup> In random assignment experiments, clients are assigned randomly to either a control or an experimental group (or non-experimental treatment group). The experimental group receives the services (the reform program) being evaluated, while the control group receives the status quo and is excluded from the reform program for a specified period of time, often on the order of three years.

<sup>2</sup> Several recent national demonstrations found slight or negative impact, and the prescribed model inherent in the design of each demonstration may have negatively affected impact results. The demonstrations in these cases can put an entire range of services to the studied group on the defensive when results are inevitably summarized into a conclusion that services to the group may not be worth an investment of funds.

<sup>3</sup> Greater Avenues to Independence (GAIN).

## **Bibliography**

Blank, Rebecca M. "Changing Policy: America's Efforts to Provide a Social Safety Net," in *It Takes a Nation: A New Agenda for Fighting Poverty*. Princeton University Press: Fall 1996.

Cain, Glen G. 1997. "Controlled Experiments in Evaluating the New Welfare Programs." IRP Special Report: Evaluating Comprehensive State Welfare Reforms: A Conference. University of Wisconsin, Madison: Institute for Research on Poverty.

Corbett, Thomas. 1997. "The Next Generation of Welfare Reforms: An Assessment of the Evaluation Challenge." IRP Special Report: Evaluating Comprehensive State Welfare Reforms: A Conference. University of Wisconsin, Madison: Institute for Research on Poverty.

Doolittle, Fred and Linda Traeger. 1990. *Implementing the National JTPA Study*. New York: Manpower Demonstration Research Corporation.

Doolittle, Fred and Suzanne Lynn. 1998. *Working with Low-Income Cases: Lessons for the Child Support Enforcement System from Parents' Fair Share*. New York: Manpower Demonstration Research Corporation.

Edin, Kathryn and Laura Lein. 1997. *Making Ends Meet: How Single Mothers Survive Welfare and Low-Wage Work*. New York: Russell Sage Foundation.

Gordon, Anne, Jonathan Jacobson and Thomas Fraker. 1996. *Approaches to Evaluating Welfare Reform: Lessons from Five State Demonstrations*. Princeton: Mathematica Policy Research, Inc.

Heckman, James J. 1992. "Randomization and Social Policy Evaluation." In Manski, Charles F. and Irwin Garfinkel. 1992. *Evaluating Welfare and Training Programs*. Cambridge: Harvard University Press, pp. 201 - 230.

Hotz, V. Joseph. 1992. "Designing an Evaluation of the Job Training Partnership Act." In Manski, Charles F. and Irwin Garfinkel. 1992. *Evaluating Welfare and Training Programs*. Cambridge: Harvard University Press, pp. 76-114.

IRP Special Report: Evaluating Comprehensive State Welfare Reforms: A Conference. 1997. University of Wisconsin, Madison: Institute for Research on Poverty.

Manski, Charles F. 1995. "Learning about Social Programs from Experiments with Random Assignment of Treatments." University of Wisconsin, Madison: Institute for Research on Poverty. Discussion Paper No. 1061-95.

Manski, Charles F. and Irwin Garfinkel. 1992. Evaluating Welfare and Training Programs. Cambridge: Harvard University Press.

Mayer, Susan E. 1997. "Has America's Anti-Poverty Effort Failed?" Northwestern University: Institute for Policy Research.  
[\[http://www.library.nwu.edu/publications/nupr/mayer.html\]](http://www.library.nwu.edu/publications/nupr/mayer.html)

Piliavan, Irving and Mark Courtney. 1997. "Interstate Comparison of Welfare Reform Programs." IRP Focus: Evaluating Comprehensive State Welfare Reforms. University of Wisconsin, Madison: Institute for Research on Poverty. Vol. 18, No. 3, pp. 29-32

Riccio, James, Daniel Freedlander and Stephen Freedman. 1994. GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program. New York: Manpower Demonstration Research Corporation.

Stack, Carol B. 1987. "A Critique of Method in the Assessment of Policy Impact." Research in Social Problems and Public Policy. 4: 137-147.

Wolfe, Barbara and Thomas Corbett. 1997. "Institute for Research on Poverty: A Description." University of Wisconsin, Madison: Institute for Research on Poverty.