The Project on Devolution and Urban Change

Assessing the Impact of Welfare Reform on Urban Communities: The Urban Change Project and Methodological Considerations

Charles Michalopoulos
Johannes M. Bos
Robert Lalonde
Nandita Verma

MDRC

Manpower Demonstration Research Corporation

November 2000

Funders of the Project on Devolution and Urban Change

Charles Stewart Mott Foundation
W. K. Kellogg Foundation
Ford Foundation
The Robert Wood Johnson Foundation
The Pew Charitable Trusts
The George Gund Foundation
The Joyce Foundation
The Cleveland Foundation
The James Irvine Foundation
The California Wellness Foundation
The Edna McConnell Clark Foundation
U.S. Department of Health and Human Services
U.S. Department of Agriculture

Dissemination of MDRC's work is also supported by MDRC's Public Policy Outreach funders: the Ford Foundation, the Ambrose Monell Foundation, the Alcoa Foundation, and the James Irvine Foundation. In addition, the following organizations support MDRC's expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: the Arco Foundation, Grable Foundation, Ewing Marion Kauffman Foundation, Open Society Institute, and Union Carbide Foundation.

Contents

For a list of our publications, see our Web site: www.mdrc.org.

List of Tables and Figures

Acknowledgments

The ideas presented in this paper build on the work of many people other than the authors. In the infancy of the Urban Change project, there was much discussion of how impacts could be measured. In particular, a paper by Robert Moffitt and Burt Barnow provided the germ of the analytical approach discussed in the paper. Comments by members of MDRC's income studies committee also strengthened the analytical plan. Understanding the policies that were implemented in the four Urban Change sites would not have been possible without the effort of implementation researchers at MDRC, including Janet Quint, Barbara Fink, Olis Simmons, Mary Valmont, and Maria Buck. The empirical investigation could not have happened without data for Cuyahoga County assembled by Edward Wang at the Center for Urban Poverty and Social Change at Case Western Reserve University. In addition, Claudia Coulton, co-director of the Center for Urban Poverty and Social Change, provided invaluable information regarding the data and welfare reform in Cuyahoga County. At MDRC, Jim Kemple provided invaluable advice regarding the evaluation of Project Independence, and Ben Schall and Ebonya Washington provided superb research assistance in working with data from Cleveland, Los Angeles, and Florida. Robert Weber edited the paper, and Stephanie Cowell did the word processing.

The Authors

Over the past few years, our nation's social safety net has undergone a period of rapid and dramatic change, the third such period in sixty years. The last two upheavals, in the 1930s and the 1960s, left enduring legacies — Social Security, Medicare, and Medicaid are among the best-known examples — and generated far-reaching changes in American society. On August 22, 1996, President Clinton signed the historic Temporary Assistance for Needy Families (TANF) block grant legislation that heralds another period of major change. The reform embodies an interesting set of contrasts. On the one hand, it grants the states an extraordinary amount of flexibility to design whatever mix of services and benefits they think will reduce dependency and provide for the needy. On the other hand, the law is quite prescriptive about work participation requirements. Specifically, the 1996 law contains a five-year lifetime limit on receipt of federally funded cash assistance; authorizes states to impose shorter time limits at their discretion; requires that, by the year 2002, 50 percent of all recipients who have received cash aid for two years work at least 30 hours a week in order to continue receiving benefits (only about 14 percent of all recipients meet this definition currently); denies Food Stamp assistance and, at state discretion, cash benefits to all legal immigrants; allows states to withdraw 20 percent of the current state share of Aid to Families with Dependent Children (AFDC) costs; and permits 30 percent of all federal funds to be diverted to child care and other programs for non-AFDC recipients.

The Manpower Demonstration Research Corporation (MDRC) has undertaken an initiative that is examining how states, urban counties, and large cities restructure social welfare programs over the next several years. The overarching goal of its Project on Devolution and Urban Change is to assess whether "devolution" is generating the dramatic changes in policy and programs predicted by supporters and critics; just as important is to understand what difference these policies are making in the lives of low-income families. Begun in early 1997, the project will last approximately five years.

Welfare dependency and poverty have become increasingly concentrated in America's largest cities. Furthermore, welfare agencies in large cities usually have had little success in moving recipients into jobs, so that the 1996 legislation poses significant implementation challenges in these settings. For these reasons, the Urban Change project is taking place in four large urban counties: Philadelphia, Los Angeles, Miami-Dade, and Cuyahoga (Ohio), which includes the city of Cleveland.

To provide depth and breadth to the analysis of welfare's changes, the study includes five major components:

1. An *impact study* to measure the economic effects of the new policies on individual welfare recipients and potential welfare recipients

2. An *implementation study* to describe the new programs and policies as they are put in place

3. An *ethnographic study* to look in depth at the experiences of a small number of families

4. An *institutional study* to examine how the new policies and funding mechanisms affect both for-profit institutions and nonprofit and public service delivery systems

5. A *neighborhood indicators study* to assess changes in the social and economic vitality of neighborhoods

The purpose of the Urban Change project is twofold: to *describe* income sources, income amounts, and well-being after devolution; and to *infer the impacts* of devolution on individuals, communities, and institutions by estimating what would have happened under AFDC. Although the five study components are designed to build on one another, and all five will be used in determining the effects of recent reform, this design paper focuses on the first component, the individual-level impact analysis. Information for the individual impact component will come from two sources. Administrative data on AFDC, TANF, Food Stamps, and earnings will be collected for all nonelderly parents of minor children who ever received Food Stamps or AFDC/TANF between 1992 and 2002 in the four counties. In addition, 2,000 single mothers receiving welfare in high-poverty neighborhoods will be surveyed in each county's principal city. While the administrative data will allow us to precisely estimate changes in welfare and employment over time, the surveys will provide crucial information on other sources of income, child and parent well-being, family formation, experiences with the welfare system, and attitudes toward welfare and work.

This paper focuses on the portion of the impact analysis that uses administrative data. To infer impacts of the new policy on employment, earnings, and welfare, the study is using a multiple cohort design. In this design, a cohort, or group, of those receiving welfare or at risk of receiving welfare will be followed over time, and their outcomes will be compared. If patterns of behavior for cohorts who pass through reform differ markedly from patterns for those who are subject to the defunct rules of AFDC, this will be taken as evidence that welfare reform has had an impact. The remainder of this paper describes the multiple cohort design, investigates the power of the technique using data from a variety of sources, and discusses analytical issues which remain to be addressed.

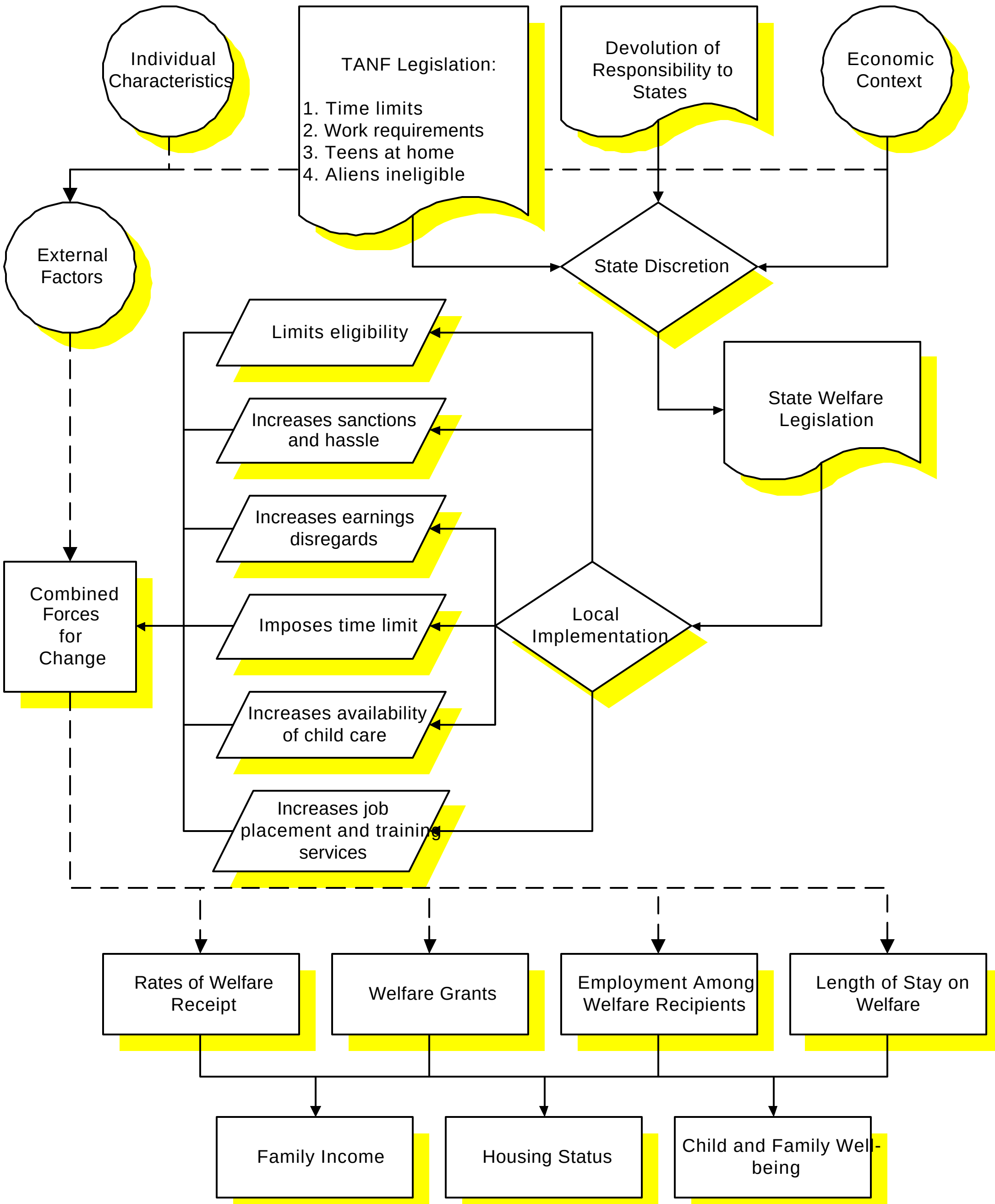
I. A Model of the Impacts of Welfare Reform: The Treatments and Predicted Effects

The likely impacts of welfare reform on state welfare agencies and on the working and welfare poor are unclear. Figure 1 provides a conceptual framework for understanding the process by which welfare reform is expected to affect individual welfare recipients, potential welfare recipients, and their families and children. The two boxes and two circles at the top of the figure describe the primary determinants of individual outcomes and changes in individual behavior. The middle boxes capture the two primary components of the welfare reform bill: (1) the TANF legislation, including the specific rules and regulations embedded in it, and (2) the devolution to the states of responsibility for design and control of cash welfare programs.

These two welfare reform boxes are separated in the flowchart to underscore the ambiguity inherent in the welfare reform legislation, which constitutes the principal reason why its impacts are difficult to predict. First, the legislation was billed as an effort to curtail and regulate welfare receipt. The provisions regarding time limits and work requirements are consistent with this emphasis. However, the legislation also transferred power and discretion from the federal government to the states. In practice, increased state discretion sometimes offsets the effects of the more restrictive parts of the legislation. For example, some states have developed programs to extend benefits to those who become ineligible under strict TANF rules.

Figure 1

A Model of the Impacts of Welfare Reform on Individuals



To the sides of the two welfare reform boxes at the top of Figure 1 are two circles depicting external factors: the economic context and individual characteristics. Changes in these external forces affect not only what the ultimate outcomes will look like (as indicated by the bold dashed line running across and then down the left side of the figure) but presumably also what states will do with their newfound freedom to design welfare programs.

The next steps in the flowchart are the design and implementation of state welfare programs reflecting TANF rules, state policymaking, and economic constraints. As with the TANF legislation, it cannot be assumed that whatever is written down in state plans and state law adequately defines the new welfare environment in the cities under study. In the model, a diamond (local implementation) separates state legislation from the actual "treatments" affecting current and future welfare recipients. This highlights the importance of the implementation study both for the Urban Change project generally and for the impact analysis in particular. If the implementation study indicates that welfare reform is not associated with real changes at the level of local welfare offices, then the impact analysis cannot be expected to detect much about the effects of the new legislation.

The arrows lead from the implementation diamond at the center of Figure 1 to the heart of the impact evaluation of welfare reform: the definition of the welfare reform treatments, which include the actual changes in eligibility rules, benefits, and obligations that will directly affect people's outcomes and behavior. The model shows six of the key changes that have been implemented, namely, (1) limits on eligibility to receive cash welfare, (2) increased sanctioning to induce welfare recipients to find work or participate in required activities and increased hassle to discourage individuals from applying for welfare, (3) increased earnings disregards to stimulate employment, (4) time limits, (5) increased resources for child care to help welfare recipients pay for one of the key work-related expenses, and (6) increased resources for job placement and training services to help welfare recipients find jobs and gain skills needed for employment. Together with the individual characteristics of the people affected and the economic context in which they live, these six treatment changes are expected to shape changes in outcomes and individual behavior. As discussed below, the key challenge in developing a viable research design is to find one that allows the effects of welfare reform to be isolated from demographic changes in the caseload and from changes in the local economy.

The bottom part of Figure 1 lists some outcomes associated with welfare reform and divides them into two levels. The first level includes outcomes on which welfare reform should have more direct or immediate effects. These outcomes, which can be measured using large samples from administrative data, include caseloads, welfare grants, employment levels among welfare recipients, and lengths of stay on welfare. The second level of outcomes is affected less directly and will be measured using smaller samples from survey data. These include family income, housing status, and other measures of child and family well-being.

The legislation passed in the four states of the study includes several major changes in welfare (see Table 1). Perhaps the most publicized change is a limit on the number of months that a parent can receive welfare. The federal legislation limits lifetime receipt of TANF benefits to 60 months, although each state is allowed to exempt 20 percent of cases from this lifetime limit. Of our four states, only Florida has set a lower limit on lifetime receipt, 48 months. In addition to the lifetime limit, three of the four states in the Urban Change project limit the number of consecutive months that welfare can be received. In California, a current recipient must leave the welfare rolls for at least one month after

Table 1

Key Features of Welfare Reform Legislation in Urban Change Sites

Key Features	Cleveland[a] Ohio Works First	Los Angeles[b,c] CalWORKs	Miami[d] WAGES	Philadelphia[e] Act 35
Administration	county-administered	county-administered	state-administered	state-administered
Time limits	3 years; lifetime limit of 5 years	18 or 24 months; lifetime limit of 5 years	24 of 60 months or 36 of 72; 48-month lifetime limit	5-year lifetime limit
Work requirements (single parent)	30 hours per week	20 hours per week; will be increased after July 1998	40 hours per week	20 hours per week (after 24 months)
Family cap	no	yes	for first additional child, clients receive half the previous benefit increase; no further increases	no
Income disregards	first $250 + ½ remainder of earned income for 18 months	first $225 + ½ remainder of earned income	first $200 + ½ remainder of earned income	½ of earned income
Employment services	job search and preparation, prevention and retention services, Work Experience Program	job search readiness and other employment services for up to 18 months, community service jobs	independent job search, job club, vocational skills training	job training and readiness services
Child support pass-through	eliminated to fund a 6% increase in TANF benefits	$50 disregard continued	eliminated	eliminated
Child care	subsidies available to all TANF recipients	eligibility set at 75% of state median income	recipients not guaranteed child care	additional $52 million invested
Teenage mothers	required to live at home	required to live at home	required to live at home; benefits to alternative payee	required to live at home
Immigrants	full eligibility for TANF and Medicaid for those who arrived before Aug. 22, 1996	full eligibility for TANF and Medicaid for those who arrived before Aug. 22, 1996	full eligibility for TANF and Medicaid for those who arrived before Aug. 22, 1996	full eligibility for TANF and Medicaid for those who arrived before Aug. 22, 1996

SOURCES: [a] "What Welfare Reform Will Mean to the People of the State of Ohio" (Ohio Department of Human Services, July 1997).
[b] "Welfare Reform Becomes a Reality" (State of California Web site, January 1997).
[c] "California Welfare Reform: Summary of the Provisions of AB 1542 and AB1008" (County Welfare Directors Association of California, August 1997).
[d] "A Plan to Reform Welfare in Florida WAGES: Work and Gain Economic Self-Sufficiency" (Florida State Legislature, May 1996).
[e] "TANF State Plan" (Pennsylvania Department of Public Welfare, enacted March 1997).

receiving benefits for 24 consecutive months; an applicant must leave the rolls after receiving benefits for 18 consecutive months. In Ohio, welfare can be received for 36 months in any five-year period. In Florida, a current recipient can receive welfare for 36 months in a six-year period, while a new applicant is restricted to 24 months of receipt in a five-year period.

The second major change in policy affects earnings disregards, allowing welfare recipients to keep more of their earnings. Under the rules of AFDC, a working welfare recipient was allowed to keep the first $120 of monthly earnings plus one-third of remaining earnings for the first four months of employment. For the next eight months, the working welfare recipient could keep $120 of earnings per month before having her grant reduced. After a year of working, the recipient could keep only $90 of earnings. Under the new rules in Florida, California, and Ohio, the fixed portion of the earnings disregard has been increased to $200 or more. In all four states, a TANF recipient can keep one-half of remaining earnings after the fixed disregard. In addition, in Florida, California, and Pennsylvania, the enhanced disregard is available no matter how long the welfare recipient combines work and welfare; in Ohio, it is available for 18 months.

The third major change in policy requires recipients to engage in work activities. One parent in each two-parent household must be in work activities for at least 35 hours per week. For single-parent cases, hours of required work were 20 hours per week in fiscal year 1997 and are 30 hours per week in fiscal year 2000. States must gradually increase the percentage of recipients meeting this work requirement; 25 percent of single-parent recipients were supposed to meet the work requirement in fiscal year 1997, but 50 percent are supposed to meet it in fiscal year 2002; of two-parent cases, 75 percent were supposed to meet the work requirement in fiscal year 1997, but 90 percent are supposed to meet it in fiscal year 1999 and later.

In addition to these changes in major treatments, welfare rules have changed in other ways. In Philadelphia, recipients are allowed to own one car of unlimited value, whereas Florida's recipients are allowed to own a car worth no more than $8,500. In Ohio, Florida, and Pennsylvania, the $50 child support pass-through was eliminated. In California and Florida, a family cap limits the increase in benefits for children added to a family after the family starts receiving TANF benefits. Some states will provide additional transitional benefits to encourage recipients to leave welfare for work.

The combined effects of these changes is less than clear. Consider the effects of the enhanced earnings disregards and increased child care assistance. Both components of welfare reform make work more attractive to current welfare recipients. Among recipients who are not currently working, this should lead to increased hours of work. However, enhanced disregards will encourage work only among people who can remain eligible for welfare while working. For this group, then, the two policies will increase work by a limited amount but will not reduce welfare use immediately.

Among recipients who already work, the effects of earnings disregards and child care assistance on hours worked are ambiguous. Both policy changes allow recipients to keep more of each additional dollar earned, and this provides an incentive to work more. Because they also allow an individual to maintain the same material living standard and to work less, however, they could cause hours of work to decline. In any case, because welfare is more compatible with work, these welfare recipients are expected to stay on welfare longer. Such behavior would have the effect of increasing the caseload.

For people not currently receiving welfare, earnings disregards and child care assistance make welfare more attractive. In the absence of some countervailing condition that makes welfare unattractive, this group's welfare use should increase. Further, because of the availability of welfare, their hours worked should decline.

Because of the combined effects of earnings disregards and child care assistance for the different groups, it is clear that the welfare caseload is likely to increase and that employment is likely to increase, but hours worked could either rise or fall. Total hours worked could decline even among recipients, although that is very unlikely. Over the longer term, the effects of these policies are even less clear. Although enhanced disregards and child care assistance will initially increase welfare caseloads, the work experience acquired could lead to higher-paying employment opportunities, which would entail both greater work effort and a lower caseload.

The presence of time limits does not alter the foregoing analysis or diminish the ambiguity associated with these components of welfare reform. Rather, it increases the uncertainty about the overall effects of welfare reform on some outcomes. For some individuals, time limits might make welfare less valuable and thus discourage welfare use. For these individuals, hours worked are expected to increase. For many potential recipients, however, the time limits do not constitute a constraint on behavior, and policies such as enhanced disregards make welfare a more attractive form of temporary assistance, increasing welfare use and perhaps lowering hours worked. Overall, then, time limits should increase the effects of welfare reform on employment, make the effects on welfare caseloads unclear, and leave the effects on hours worked ambiguous.

The ambiguities associated with the effects of the likely components of welfare reform motivates the Urban Change impact study. Because theory provides little guidance about the impacts of reform, only empirical analysis can shed light on the study's key questions. Both positive and negative impacts may occur, leaving open the question of the magnitude of these effects and the net impact of TANF on individuals, their communities, and public sector budgets.

II. The Proposed Design: A Multiple Cohort Analysis

The process of measuring the effects of welfare reform on individual outcomes has three distinct steps. The first step is to measure the outcomes of interest for those affected by the changing welfare environment. The second step is to identify a counterfactual state for the changing welfare environment and to measure outcomes for this counterfactual state. The third step is to compare outcomes for the changed environment and its counterfactual, thereby calculating the *impact* of the change. Expressed as a simple equation, this looks as follows:

Impact Welfare Change = (Outcome|Welfare Change) – (Outcome|No Welfare Change)

The last term in this relation is the *counterfactual*. Because it is easy to measure outcomes for those affected by the policy change, the search for an appropriate research design centers on finding the best method of estimating the counterfactual.

Alternative I: Random Assignment

The research design most commonly used by MDRC to obtain the counterfactual outcome is random assignment of some individuals to a control group that is denied access to the program being studied. As is widely understood, the randomness of the procedure by which a control group is created protects the estimated program impacts from virtually all potential sources of bias. However, random assignment has several important drawbacks. First, it is expensive and difficult to implement. Second, it often raises ethical questions. The third and most important drawback is that random assignment is useful only in situations in which the control group represents the intended counterfactual. Arguably, the changes in U.S. welfare policy since 1996 do not constitute such a situation. Although it is theoretically possible to randomly assign welfare recipients to a traditional AFDC program kept in place for research purposes only it would be difficult to isolate traditional AFDC control group members from the system-wide and community-wide changes going on around them. Therefore, a random assignment design does not seem a feasible option for this evaluation of welfare reform.

Alternative II: Cross-Sectional Comparisons

A related nonexperimental method of evaluating a new policy is to compare outcomes for different individuals, some of whom are subject to welfare reform, some of whom are not. This type of approach requires constructing or finding a *comparison group* that is the nonexperimental equivalent of the control group used in experimental evaluations. For a comparison group to provide an appropriate counterfactual, two conditions must be met, namely, (1) that the comparison group not be subject to a treatment and (2) that the comparison group be comparable to the program group in all aspects relevant to both the treatment and the outcome.

In evaluating the effects of TANF, neither requirement can be met. The global nature of the welfare changes makes it unlikely that a relevant comparison group can be found that is not itself affected by welfare reform. Nonexperimental evaluations of other policy changes have sometimes used variation in programs across states or counties to isolate a program's effects. However, all states were affected by the TANF legislation, and even though there is variation in welfare programs across the states, such variation is probably insufficient to produce a compelling counterfactual for welfare reform itself.

Alternative III: Time-Series Analysis

Many policy changes have been evaluated by examining changes in outcomes over time. There are many different types of such time-series designs, but ultimately they all use the past as the primary source from which to create a counterfactual to measure the effects of an intervention. Time-series analyses are popular, because they are intuitively appealing. They also offer a very compelling source of causal inference by imposing a temporal order on otherwise ambiguous relationships among variables.

The simplest example of a time-series design is known as a *before-after analysis*, which is easy to understand and quite popular with journalists, politicians, and voters. This design simply compares an outcome measured before a change takes effect with the same outcome measured after the change. A hypothetical example of a before-after analysis would be a comparison of Los Angeles County's welfare caseloads in 1994 and 2000. Such a design would not be considered for the Urban Change project because of the many problems inherent in such a comparison. These problems manifest themselves as three different threats to the internal validity of the analysis, namely, (1) maturation, (2) history, and (3) regression to the mean (Cook and Campbell, 1979). Following is a brief discussion of these validity threats.

Maturation. Maturation bias is the most obvious threat to validity and also the easiest to eliminate. Maturation represents the idea that outcomes may gradually change over time as a result of natural processes. For example, any welfare recipient will, over time, leave the rolls as her youngest child reaches the age of 18. Or, in another example, a city's rate of welfare receipt may gradually increase as middle-class households move away to suburbs outside the city limits.

Maturation bias in a two-point time-series design is illustrated in Figure 2, in which time is represented on the horizontal axis while a generic outcome Y is represented on the vertical axis. The researcher observes outcomes only at two points — point A, showing the outcome prior to the policy change, and point B, showing the outcome after the policy change. The dotted line shows an unobserved underlying trend in Y. Without knowing the trend, a researcher might attribute the change in Y between points A and B to the program, estimating its effect on Y to be b-a. However, the trend shows that this conclusion is wrong. Pre-program measure A is not a good counterfactual for post-program measure B. Instead, the appropriate comparison would be between points C and B, producing an estimated effect of b-c, which is the exact opposite of the original estimate. To address this problem, the researcher needs to add data about point D to the analysis, thereby allowing one to predict the location of point C (by extrapolating the line DA) and forming an appropriate counterfactual for post-program outcome B. This means that, given a stable trend in Y, all that is needed to prevent maturation bias is to find one or more pre-program points D to obtain a good estimate of the counterfactual C.

History. A more difficult problem is created by the second validity threat, history. In studying the effects of a policy change, the problem of history occurs when a second, unrelated change occurs after the new policy goes into effect. Figure 3 provides an example.

Assume that there are sufficient pre-program data to estimate a trend for outcome Y. That is, assume that the researcher can observe point D and point A and, by extrapolation of line DA, that point C is chosen as the counterfactual. To this researcher, b-c appears to be the impact of the new program. Unknown to the researcher, however, a second event has occurred that affects outcome Y. An example of this event could be the enhancement of the local JOBS (Job Opportunities and Basic Skills Training) program or a state-mandated reduction in welfare grants, unrelated to the new TANF legislation. In the example shown in Figure 3, the event changed the direction of the underlying (unobserved) trend line, directing it downward, toward point C. This means that the true impact of the second event (welfare reform) is b-c' instead of b-c. The difference is attributable to history bias.

The extent of history problems in time-series analyses is determined by a number of different factors, including (1) the number and spacing of individual observations of Y, (2) the ability to identify the event causing the bias, (3) the ability to separate the timing of different events, and (4) the size of the spurious effect on Y relative to the program effect. In other words, a time-series design is least likely to suffer from history bias if there are observations at many points in time, if important events are widely spaced, if there are data for all relevant time-varying variables, if changes are sudden and timing is precise, and if program effects are large compared with anything else that may occur simultaneously.

Figure 2

Maturation

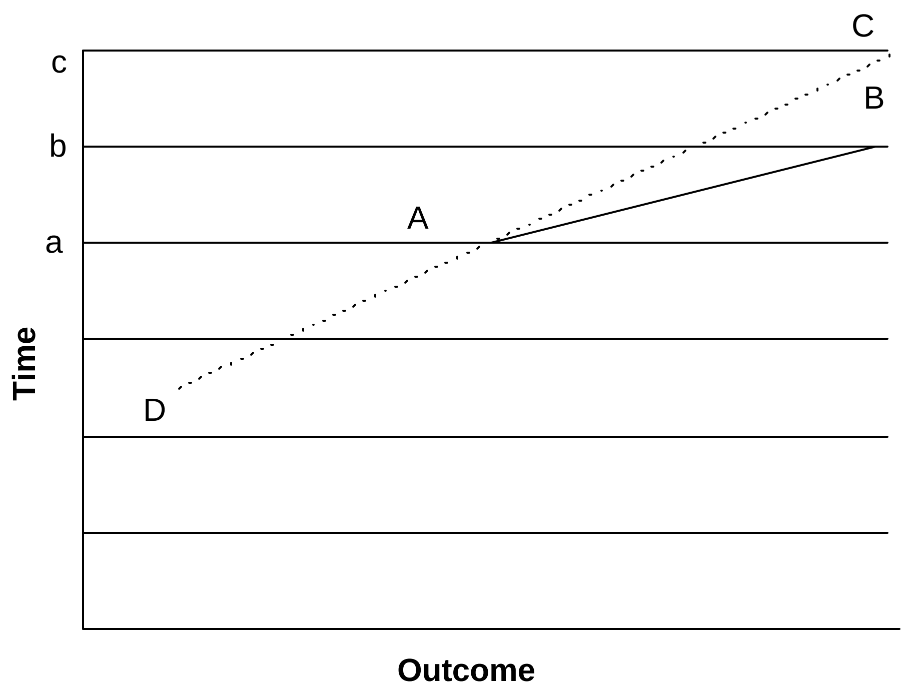
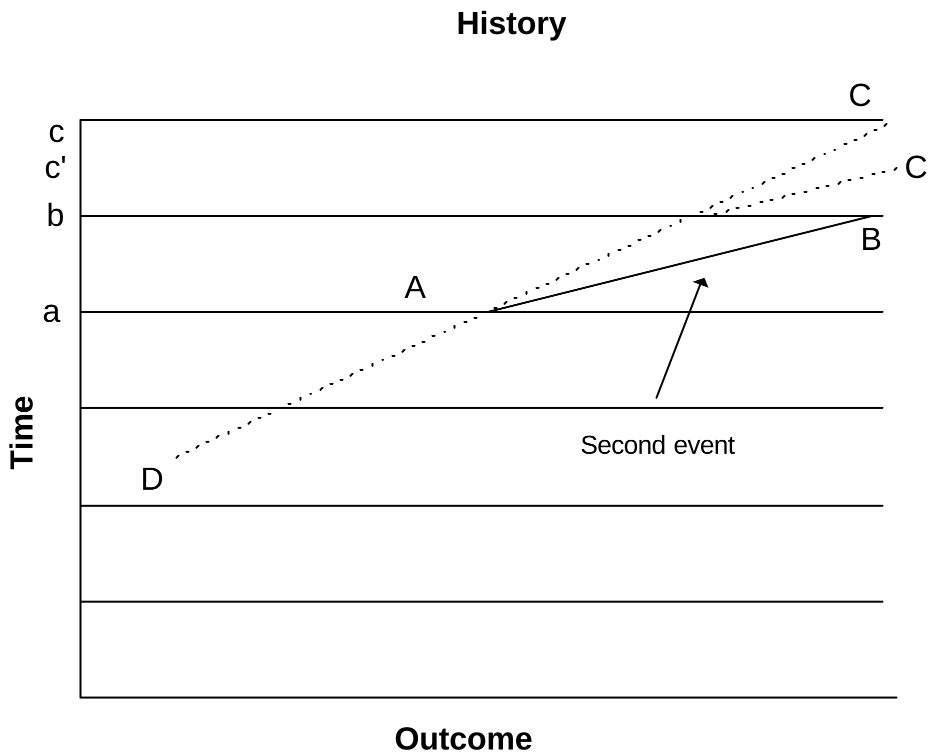


Figure 3



(The discussion below about more developed research designs will return to these criteria.)

Regression to the mean. This third type of bias prevalent in time-series designs is a special type of maturation bias that is much more difficult to control with pre-program data because the underlying trend is not linear. Consider Figure 4. The dotted line shown in earlier figures has been replaced by a slow-moving "wave" that is centered on an underlying trend line (not shown). From each extreme point of the wave, the outcome has a tendency to move back toward the underlying mean. This movement reflects a "natural" tendency and is considered unrelated to the underlying trend and also unrelated to any events happening along the way.

In many situations, regression to the mean is a simple source of random variation, which does not cause estimates to be systematically biased. However, there are situations in which the selection of a sample or its exposure to a treatment systematically occurs at the same time as an extreme in the oscillating trend. Applying to receive welfare is a good example of such a situation. At the time of application, people are likely to have reached a natural low in their earnings and other resources; the implication is that those outcomes will likely increase by themselves without any additional intervention. Such a low is often referred to as the "pre-program dip," and it is a serious problem in many evaluations that rely on time-series designs.

Figure 4 shows an extreme example of the potential bias from regression to the mean. Using points D and A to estimate the underlying trend, a counterfactual is created at point C. However, the real trend follows the dotted line (rather than the dashed one), leading to the real unobserved counterfactual at point C'. The bias is the sum of bc' and cb.

The only way to prevent a bias such as this one (while still remaining within a time-series framework) is to collect enough pre-program data to reliably estimate the underlying trend line. If the periodic ups and downs in Y are spaced widely apart, however, this means going back in time several years at least. Such history may not always be available.[1]

Alternative IV: Combinations of Cross-Sections Over Time (Cohort Analysis)

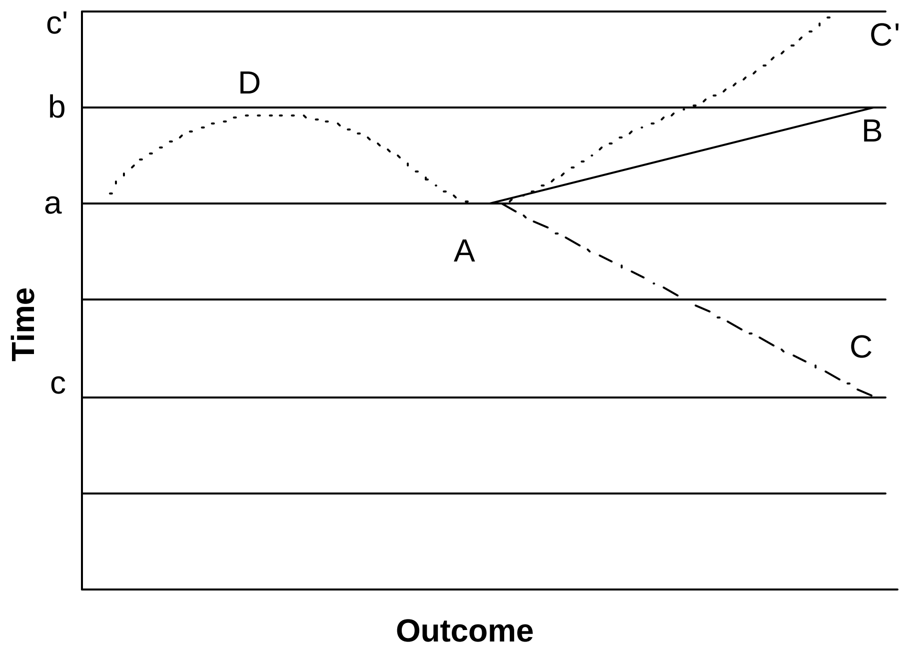
An approach that combines the advantages of cross-sectional analysis with the advantages of time-series analysis would use information about several cross-sections chosen at different points in time. An analysis using this type of approach is commonly referred to as *cohort analysis*, because it compares different cohorts of people drawn from essentially the same population over time. Cohort analyses have been popular in situations where there is frequent turnover of the population of interest, for example, in the evaluation of school reform programs, where the same grade can be compared over multiple school years.

In a cohort analysis, some of the cross-sections are selected from data collected before welfare reform, and some are selected from data collected after welfare reform. The cohorts selected before welfare reform provide comparison groups that have the advantage of not having

[1]Note that on an aggregate level, regression to the mean can also be a serious problem. As particular social problems (such as crime) reach periodic peaks, new policies to combat them may gain support and be put into place. The social problem subsequently declines, leaving policymakers and politicians proud of their achievement. Next, interest in the social problem weakens, funding goes down, and the social problem begins a new (natural) upswing, generating calls for increased funding and new programs once again.

Figure 4

Regression to the Mean



been subject to welfare reform (at least at the point of selection). Comparability of different cohorts could be achieved by choosing similar individuals at different points in time — provided that the population of interest does not change dramatically over time.

While combining the benefits of the time-series approach and the cross-sectional approach, the cohort analysis addresses important drawbacks in each. Compared with a straight time-series analysis, a cohort analysis is more robust in withstanding biases from maturation and regression to the mean, because subsequent cohorts being compared are similarly selected and thus are at similar points on their trend lines. Compared with a straight cross-sectional approach, an earlier cohort functions as a powerful and intuitively appealing comparison group, which by construction originates in the same population.[2]

Even though cohort designs address some of the problems inherent in time-series and cross-sectional analyses, some caveats remain. The most important problem concerns the threat of *history*. As in a time-series analysis, the definition of the "treatment" includes all relevant changes occurring during the time that separates the different cohorts. Therefore, when an outcome is affected by multiple changes in policy or economic circumstances that occur in a short period of time, it may be very difficult to attribute any of the changes in the outcome to one particular event or program. It is possible to alleviate this problem by drawing multiple cohorts in relatively short intervals (that is, in months or quarters instead of years), thereby introducing more variation into the analysis and producing a more comprehensive picture of developments over time.

Another potentially important problem concerns the issue of the comparability of cohorts. If people self-select into a certain population, cohorts drawn from that population will reflect the underlying selection process. To the extent that this selection process is exogenous to the program being evaluated, there is no problem; unfortunately, however, there are many situations imaginable in which the selection process would be affected by changes in the program. A clear example of such a situation would be an analysis of outcomes for pre- and post-reform cohorts of welfare recipients. If welfare reform affects who goes on welfare (through entry or deterrence effects), a comparison of outcomes for welfare recipients in different years may suffer from severe selection biases. Therefore, it is necessary either to establish conclusively that there are no entry or deterrence effects or to select cohorts from a more neutral population, whose composition is unaffected by changes in the welfare environment.[3]

The design proposed for the Urban Change impact analysis extends the idea of multiple cohorts one step further, by following each cohort over time. Thus, rather than being a combination of cross-sections over time, the multiple cohort analysis will be a combination of panels over time. The next section indicates the intuitive thinking behind this approach by offering several hypothetical examples.

[2]Ideally, one's cohort would be determined by exogenous (and essentially random) factors — like year of birth — in which case, a cohort analysis becomes like a natural experiment.

[3]It is also possible to adjust estimates for measured differences in the demographic characteristics of different welfare cohorts.

III. Hypothetical Examples of Multiple Cohort Analysis

Figure 5 is a graphical example of multiple cohort analysis that demonstrates the robustness of the design in withstanding a type of maturation bias. If this were the Urban Change impact analysis, each downward-sloping line might represent everyone who had received Food Stamps at any time during the year when the line is first drawn. The line labeled A, for example, might be the proportion receiving welfare among everyone who received AFDC or Food Stamps for any month in 1993. Likewise, lines B, C, and D would show the proportion receiving AFDC over time for everyone who received AFDC or Food Stamps in 1994, 1995, or 1996, respectively.

In this imagined case, welfare reform appears to have no effect. It is true that the 1997 cohort — defined just after TANF went into effect — receives less and less welfare over time. However, each successive cohort shows exactly the same reduction in its welfare use over time. The reduction in welfare use over time for Cohort A is just as great as for Cohort D, as well as for all other cohorts.

It is easy to extend this example to other cases in which changes over time represent ongoing trends. For example, it is possible that the recent declines in welfare caseloads have resulted because welfare recipients leave welfare faster than in the past. With the multiple cohort design, it is possible to correct for *systematic* changes in the rate at which people leave welfare. In particular, it is possible to ask (1) whether welfare spell lengths were decreasing even before welfare reform went into effect and (2) whether further decreases in welfare spell lengths after TANF represent part of the preexisting trend or an acceleration of that trend.

The multiple cohort design may also make it easier to detect and correct for history bias. Figure 6 provides an example to demonstrate this possibility. In this example, welfare receipt for each cohort declines over time. However, several other changes also appear to be happening. In this hypothetical example, at the time that welfare reform went into effect, near the end of 1996, welfare receipt for each cohort jumps up considerably. This could represent the effects of enhanced earnings disregards that were implemented in most states as part of the TANF changes. This hypothetical site has a two-year time limit, as Florida does for the most job-ready recipients. Two years after the implementation of TANF, welfare receipt in this example drops for all cohorts that existed at the time of welfare reform. This could represent the effects of cases closing as individuals hit their time limit.

One other change also appears in Figure 6: About a year before welfare reform, welfare receipt drops considerably for all cohorts. This may be an example of history bias. Without noting the change, one would expect welfare receipt to be much higher after TANF reform, which would lead to understating the effect of TANF on welfare receipt. In the multiple cohort design, however, the same change would be observed in each cohort, and one would infer that something had happened at this point to change the trajectory of welfare receipt for the various cohorts. Having multiple cohorts, one would also note that this is an unusual change that does not appear to reflect the normal patterns of welfare receipt. The change occurs three years after Cohort A is observed, two years after Cohort B, and a year after Cohort C. For all three cohorts, however, the event happens near the end of 1995. After correcting for this change, one would correctly note the large jump in welfare receipt when TANF reforms went into effect.

Figure 5

Hypothetical Example of Multiple Cohorts: Welfare Reform Has No Impact

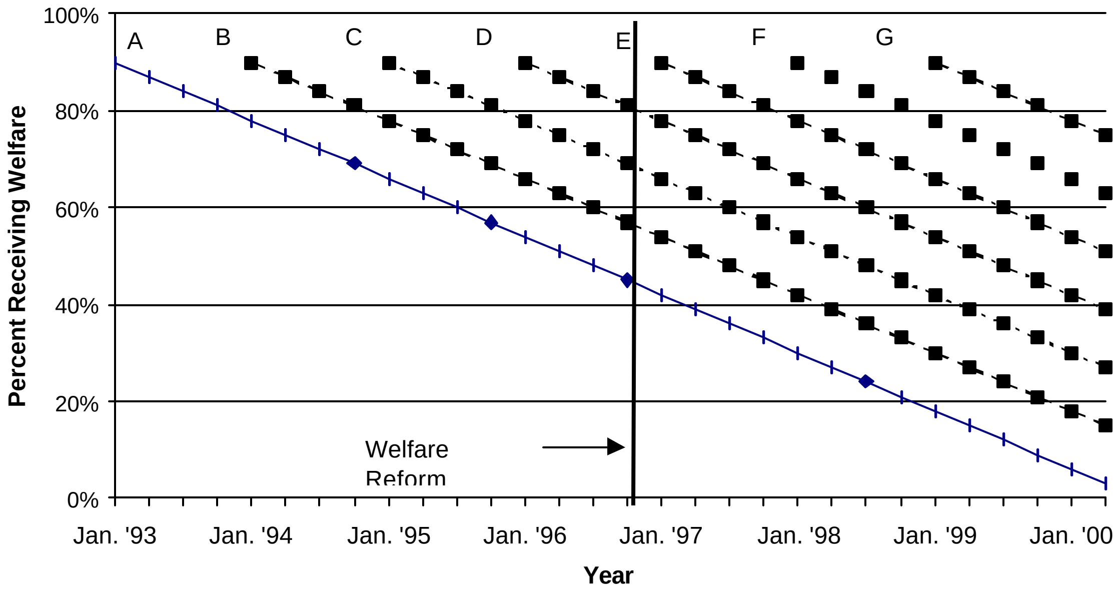
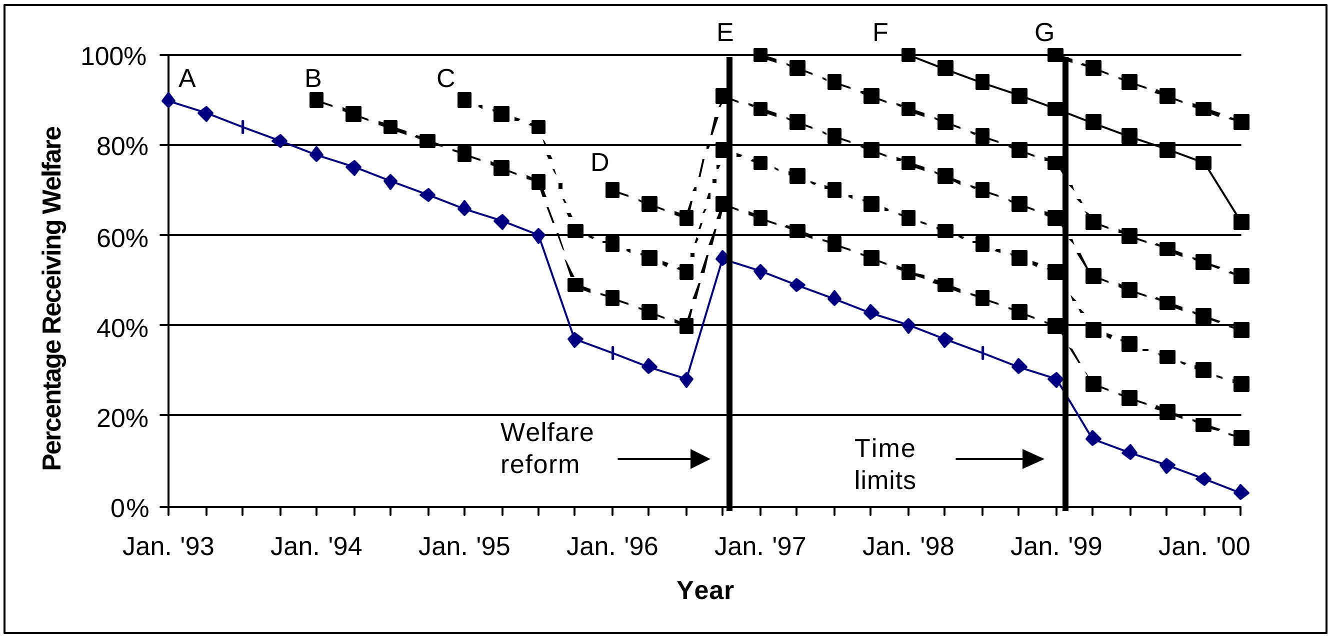


Figure 6

Hypothetical Example of Multiple Cohorts:
Initial Reform Increases Caseloads; Time Limits Reduce Caseloads



This example also highlights how important the implementation research and the other components of the Urban Change project are to the impact analysis. Changes in the example appear to have happened at three points in time — the end of 1995, the end of 1996, and the end of 1998. If such changes were to occur, implementation research would help in interpreting the changes. For example, the federal reform went into effect in October 1996, but many states did not implement reform until much later. In Los Angeles County, for example, the TANF program did not go into effect until 1998. Implementation research will help identify when the reforms were fully implemented, so that one would know whether the change that is shown for 1996 might represent the effects of new TANF policies. In Figure 6, the change shown at the end of 1998 might reflect time limits. For example, it could represent Florida's 24-month time limit. Implementation research would have revealed, however, that Florida gave extensions to most cases who hit its 24-month time limit, so that the hypothetical changes shown for late 1998 cannot represent the effects of time-limit case closures. More generally, the results of implementation research will be a key to understanding how and when policies were changing. Likewise, ethnographic interviews will be a key to understanding how welfare recipients understood and viewed the policy changes and when they did so.

If reality could be held as constant as the cohorts in these examples, very few cohorts would be needed. Only two pre-reform cohorts are needed to pick up the simple trends that are shown, and no other events appear to be changing the behavior of the cohorts' members. Figure 6 shows a more realistic hypothetical example that attempts to make clearer why many cohorts are needed to conduct a credible cohort analysis. Figure 7 demonstrates a case in which welfare receipt has both a systematic component (it is generally declining over time for each cohort) and a random component (welfare receipt for each cohort jumps around quite a bit).

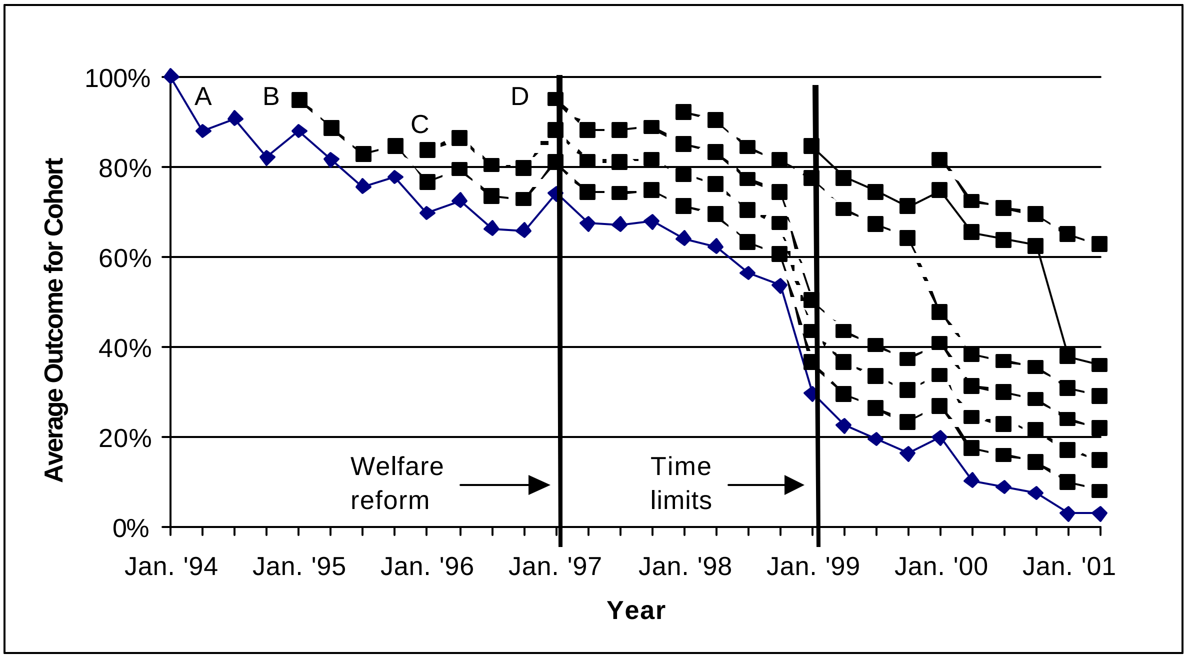
Figure 7 indicates that it might be quite difficult to infer the effects of welfare reform. As in Figure 6, welfare receipt increases considerably when welfare reform goes into effect, and it decreases considerably when time limits begin to close cases. But do these changes really reflect the policy changes, or could they represent particularly large random shocks that are similar in character to the changes that are making welfare receipt rates jump around both before and after welfare reform? The more cohorts contained in the analysis, the better the analysis will be able to separate random variation over time from trends or changes caused by welfare reform. The more narrowly spaced the cohorts, the smaller will be the variations from period to period, and the better able the analysis will be to discern changes resulting from changes in welfare policy. The next two sections use real data to investigate the amount of natural variation in several outcomes across cohorts and, by extension, to indicate how large the effects of welfare reform will have to be before one can reliably separate them from natural variation.

IV. Random Assignment Evaluation as Multiple Cohort Analysis

The simplest example of a cohort analysis is a random assignment evaluation, in which a cohort called the control group is compared with a cohort called the program group. Even the terms of a random assignment evaluation can be used in discussing multiple cohort design. The time during which a cohort is first observed is analogous to the period of random assignment. The periods after that are similar to the periods of follow-up in an experimental design. Finally, the age of a cohort is analogous to the amount of time that has passed since random assignment.

Figure 7

Percentage Receiving Welfare in Several Imagined Cohorts:
All Cohorts Are Affected by Unspecified Time-Specific Changes



To provide an example of a multiple cohort analysis, this section analyzes data from MDRC's evaluation of Florida's JOBS program, Project Independence. From July 1990 through August 1991, applicants for AFDC and AFDC recipients at the point of redetermination were randomly assigned to treatment and control groups in seven counties in Florida.[4]

If the entire randomly assigned sample of control group members and program group members from the evaluation is used, the analysis provides experimental impacts of the program. The first column of Table 2 shows results of such an analysis; that is, it shows the difference in the average outcomes for program group members and control group members. According to the experimental results, Project Independence had consistently modest impacts, decreasing AFDC receipt by 2 to 3 percentage points in most quarters of follow-up and increasing employment by no more than 2 percentage points in any quarter.

Bias in a Nonexperimental Multiple Cohort Analysis

A random assignment evaluation differs from a nonexperimental cohort analysis in one crucial respect. In a random assignment evaluation, each cohort contains both the program (treatment) group, which is subject to the new policy, and the control group, which is not. As a result, the program and control groups not only are similar in demographic characteristics and past experiences but also live in the same economic and policy environment (with the exception of the policy being tested). In contrast, in a nonexperimental analysis, at any point in time, either all members of a cohort have experienced the new policy or none of them has. As a result, impacts from a nonexperimental cohort analysis such as will come from the Urban Change project will confound the true impact of the policy change with other differences that exist between the cohorts.

Since the true impacts of Project Independence are known, data from the evaluation of that program provide an excellent opportunity to investigate how estimates of impacts are biased by changes in the economy and by changes in policy unrelated to the treatment. In the middle of the period of random assignment for the evaluation of Project Independence, the economy worsened. Between January and July 1991, state-wide unemployment rates increased from about 6 percent to about 8 percent, and welfare caseloads grew from about 150,000 to about 170,000 (Kemple, Friedlander, and Fellerath, 1995). Moreover, in the first part of 1991, the State of Florida cut staffing for Project Independence and cut subsidies for child care for the program's participants. These changes suggest that there should be differences among cohorts — higher earnings and less welfare a year after random assignment for the later cohorts if their labor market potential and history were better; lower earnings and more welfare a year later if the cut in funding made it harder for these people to make the transition from welfare to work.

To examine the bias in a nonexperimental cohort analysis, Table 2 also presents two alternative estimates of the impacts of Project Independence. The second column compares outcomes for program group members who were randomly assigned between July and December 1990 with outcomes for control group members who were randomly assigned after December 1990. The third column compares outcomes for program group members who were randomly assigned after December 1990 with

[4]For more information on Project Independence, see Kemple, Friedlander, and Fellerath, 1995.

Table 2

Estimated Impacts of Project Independence on Quarterly Rates of AFDC Receipt, AFDC Benefits, Employment Rates, and Earnings

		Alternative Cohort Estimates	
Quarter of Follow-up	Estimated Impact Using All Individuals Randomly Assigned	Estimated Impact Using Program Group Cohorts 1 and 2 and Control Group Cohorts 3 Through 5	Estimated Impact Using Control Group Cohorts 1 and 2 and Program Group Cohorts 3 Through 5
AFDC receipt (%)			
Quarter 1	-1.1	-3.4	1.3
Quarter 2	-3.2	-4.7	-1.0
Quarter 3	-4.0	-7.8	0.0
Quarter 4	-3.5	-7.6	0.7
Quarter 5	-2.5	-5.2	-1.2
Quarter 6	-2.4	-1.2	-5.9
Quarter 7	-2.3	3.3	-9.3
Quarter 8	-1.8	6.3	-9.0
Quarter 9	-1.2	5.6	-6.2
Quarter 10	-0.5	3.3	-3.7
AFDC benefits ($)			
Quarter 1	-15	46	-111
Quarter 2	-32	51	-144
Quarter 3	-36	21	-112
Quarter 4	-27	15	-119
Quarter 5	-21	46	-142
Quarter 6	-21	90	-201
Quarter 7	-20	132	-227
Quarter 8	-16	151	-216
Quarter 9	-9	143	-189
Quarter 10	0	120	-162
Employment rates (%)			
Quarter 1	2.3	2.0	2.0
Quarter 2	2.0	1.1	3.4
Quarter 3	1.3	-0.7	2.6
Quarter 4	1.7	-0.4	3.4
Quarter 5	2.0	-0.5	3.5
Quarter 6	0.4	-2.3	4.0
Quarter 7	0.8	-2.8	4.4
Quarter 8	-0.2	-3.8	2.9
Quarter 9	1.0	-2.6	2.0
Quarter 10	0.5	-2.0	0.6
Earnings ($)			
Quarter 1	33	12	53
Quarter 2	77	26	150
Quarter 3	45	-17	75
Quarter 4	46	4	100
Quarter 5	49	-16	99
Quarter 6	51	-37	178
Quarter 7	28	-76	76
Quarter 8	6	-60	111
Quarter 9	32	-75	81
Quarter 10	39	-57	112

SOURCES: Kemple et al., 1995, and calculations using data from the evaluation of Project Independence.

outcomes for control group members who were randomly assigned between July and December 1990.[5] In other words, to obtain these alternative estimates, part of the program group is compared with part of the control group. Unlike an experimental comparison, however, the alternative analyses use program group members who were randomly assigned at one point in time and control group members who were randomly assigned at a different point in time. As a result, the two groups being compared are not necessarily similar, because they were randomly assigned at different points in time.

The alternative analyses imply quite different effects of the treatment. The first alternative (the second column in Table 2) implies that Project Independence initially lowered welfare receipt by more than indicated by the experimental estimates. In the last few quarters of follow-up, however, estimates using this alternative imply that Project Independence increased welfare receipt later in the follow-up period. In contrast, the second alternative (the third column in Table 2) implies that the impacts of Project Independence are two to four times as high as the experimental estimates. Thus, both analyses produce results that differ from the experimentally estimated impacts, differing by as much as 7 to 8 percentage points in some quarters.

The last three sections of the table repeat the comparisons for three additional outcomes: quarterly AFDC benefits, quarterly employment rates, and quarterly earnings. All three comparisons tell similar stories. The first alternative analysis implies that Project Independence was detrimental, resulting in lower employment and earnings and higher welfare. The second alternative analysis implies that the program was much more successful than the experimental evidence indicates.

This discussion indicates some potential avenues for trying to reduce the bias shown in Table 2.[6] If the current state of the economy is causing the bias, then including contemporaneous measures of the economy, such as the current unemployment rate or employment levels, might ameliorate the bias. If selection is causing the bias, then history might be used as a means of correcting for cohort changes. Likewise, the unemployment rate might be used when a cohort was chosen as a measure of how desperate the case was when it was seen receiving welfare. Finally, if some individuals are affected by changes in the economy more than others, then the reform should be allowed to affect different individuals differently. Some initial attempts at incorporating these factors did not alleviate the biases shown in Table 2.

V. Natural Variation in Outcomes Across Cohorts

The results in Table 2 demonstrate that an important source of bias in a multiple cohort analysis is the natural variation in outcomes from time to time. Because the Urban Change impact analysis will use information about millions of individuals composing the universe of welfare recipients, the estimates

[5]Daniel Friedlander and Philip K. Robins conducted a systematic study of the bias involved in calculating impacts using cohorts randomly assigned at different times. While they found considerable bias compared with an experimental evaluation using cohorts randomly assigned at the same time, they found that this approach results in smaller biases than other nonexperimental comparisons. See Friedlander and Robins, 1995.

[6]Two additional analyses were conducted but are not shown. These analyses attempted to reduce the bias shown in Table 2 by allowing time-specific aggregate changes and by allowing the maturation profile to have a trend across successive cohorts. Although these changes lessened the bias in some cases, they exacerbated the bias in other cases.

will be extremely precise. Nevertheless, conclusions might be wrong if they do not take into account how much variation in employment and welfare receipt normally occurs. To investigate the importance of natural variation over time, two additional data sources will be used to include information about successive cohorts of welfare recipients who were observed when there was no obvious policy change. This exercise will provide a second means of judging how large the impacts of reform will need to be in order to be confident that there has actually been an impact.

JTPA Data

The first source of data is from an evaluation of the Job Training/Partnership Act (JTPA) of 1982. The JTPA was evaluated using a random assignment design in which welfare recipients were assigned to program and control groups between January 1988 and March 1989. To investigate the variability of earnings and employment across cohorts, three cohorts were artificially defined: cases randomly assigned in the first half of 1988, those randomly assigned in the second half of 1988, and those randomly assigned in the first quarter of 1989. Each cohort contains both program group members and control group members, so that differences between them are not the result of the treatment.

Figure 8 compares monthly earnings after treatment for the three cohorts. The solid line indicates how much higher or lower monthly earnings were for Cohort 2 compared with Cohort 1. Likewise, the dashed line indicates how much higher or lower earnings were for Cohort 3 compared with Cohort 1. Finally, comparing the two lines gives an indication of the difference in earnings between Cohorts 2 and 3. The largest difference is about $40, in month 12. This can be used as one estimate of the potential bias in a cohort analysis. Suppose that, including hundreds of thousands of welfare recipients, it is found that the average recipient in a cohort after reform earned $40 more than the average recipient in a cohort before reform. This would increase confidence in the estimate of $40. Nevertheless, the JTPA sample implies that this change could easily have resulted from normal variation in the monthly earnings of welfare recipients.

Figure 9 makes a similar comparison for monthly employment rates. Between Cohorts 1 and 3 there are substantial differences in employment rates, exceeding 4 percentage points in month 6 after treatment. Cohorts 2 and 3 show an even larger difference in month 12. If the monthly employment rates of a cohort that passed through welfare reform are about 5 percentage points higher than those of an earlier cohort, the JTPA comparison implies that this could result from normal variation in monthly employment rates.

Data for First-Time AFDC Recipients in Cleveland

For each quarter from the third quarter of 1992 through the last quarter of 1994, information was obtained about new AFDC recipients in Cleveland, one of the four cities in the Urban Change study. In each quarter, the sample adds individuals who received AFDC in the first month of that quarter but who had not received AFDC since the beginning of 1990. Then the individual's earnings and the AFDC benefits for the individual's welfare case are followed through December 1995.

This sample provides two useful benchmarks. The first benchmark is how much variation in earnings and welfare receipt occurred in Cleveland *before* welfare reform. If there were a steep decline in benefits before reform, then it could be argued that there needs to be an even steeper decline in benefits during and after reform in order to show that the reform lowered welfare receipt. If there were substantial variation from cohort to cohort in welfare receipt, then any change that occur during the

Figure 8

Variation in Monthly Earnings Post-Random Assignment for Three JTPA Cohorts
Difference from Cohort 1

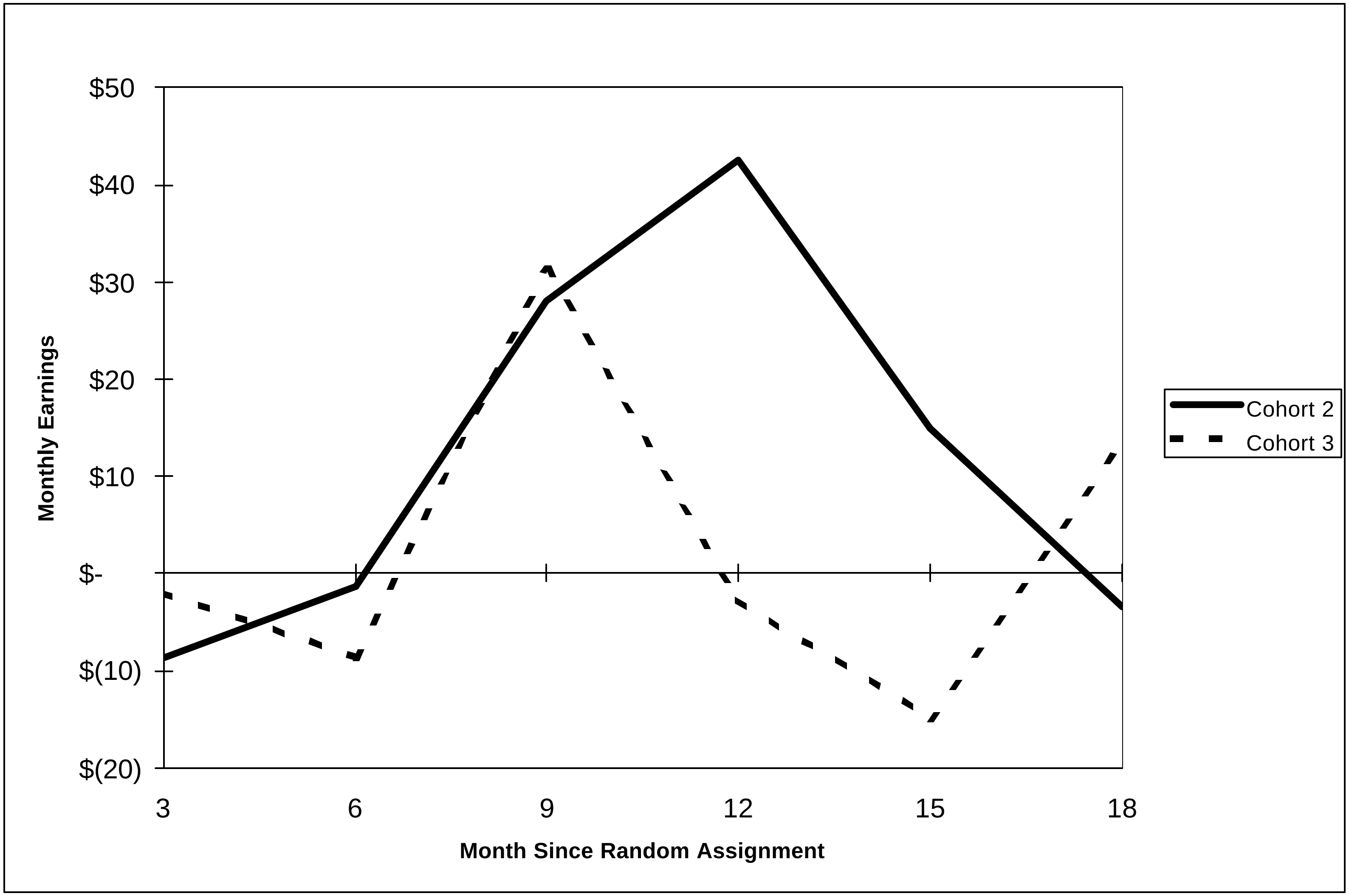
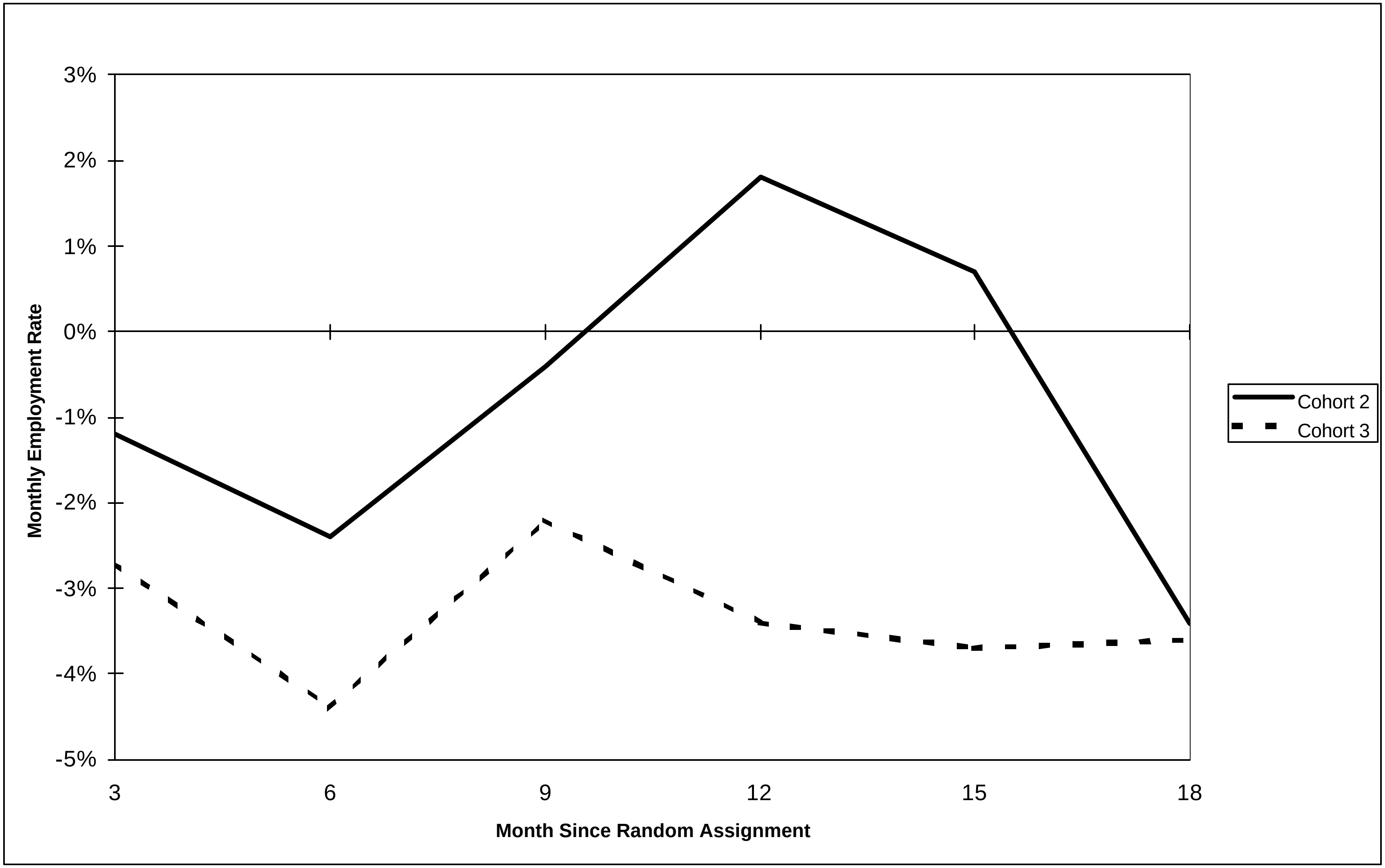


Figure 9

Variation in Employment Rates Post-Random Assignment for Three JTPA Cohorts
Difference from Cohort 1



reform need to be larger than the historical variation if the changes are to imply that welfare reform had an impact. Finally, if one performs a multiple cohort analysis using these data — artificially assigning some cohorts to a pre-reform period and some to a post-reform period — and finds an impact of this artificial and imaginary reform, that will indicate the extent of bias inherent in multiple cohort technique and will provide a benchmark which welfare reform needs to exceed in order to argue that reform had an effect.

The sample provides one with individuals who have not received welfare in at least several years and, therefore, who are likely to be new welfare recipients. As a result, the sample allows one to estimate a second benchmark that indicates how large the bias might be in cohorts of new welfare recipients.

Figures 10 through 13 show the variation from quarter to quarter in rates of AFDC receipt, AFDC benefits, employment, and earnings for each of the entry cohorts in Cleveland. In each figure, a line shows for each quarter after entry the difference between a particular cohort's outcome and the outcome for the cohort of new recipients in April 1993.

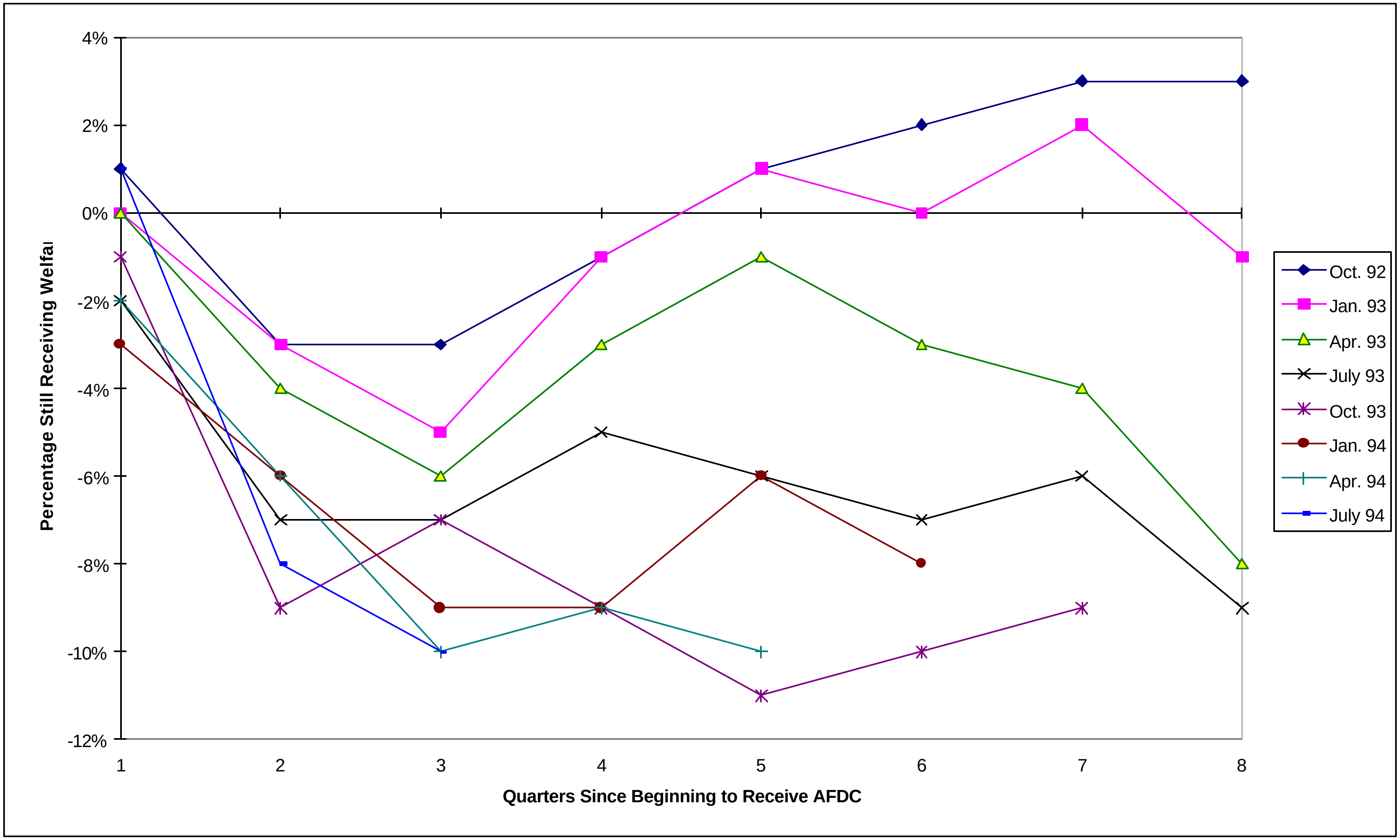
The cohorts in Cleveland show much greater variation in employment (Figure 12) and earnings (Figure 13) than did the JTPA sample. Variation in earnings is more than $400 in some quarters, while variation in employment is nearly 15 percentage points in some quarters. There is much less variation in AFDC benefits (Figure 11); benefits for the January 1994 and the October 1992 cohorts were about $50 apart. However, there is substantial variation in rates of AFDC receipt (Figure 10). These figures imply that if one found changes of 15 percentage points in employment or AFDC receipt or $500 in earnings, those changes would be well within the normal variation over time in Cleveland and, hence, could not be attributed to the welfare reform.

There are two likely reasons for the great variation in the Cleveland sample. First, the Cleveland sample contains more cohorts. Whereas all cases in the JTPA sample were randomly assigned over an 18-month period, the Cleveland sample's new recipients entered over a period of more than two years. Further, while each cohort in the JTPA diagrams (Figures 8 and 9) covered half a year, each cohort in the Cleveland diagrams (Figures 10-13) covers only three months. However, there is no clear pattern to the variation. Consider the fifth quarter after entry — the quarter with the greatest variation across cohorts. As shown in Figure 13, earnings for the January 1993 cohort are closest to those for the April 1993 cohort, but earnings for the July 1993 cohort are nearly the farthest away from the April 1993 cohort.

A second reason for the variation is that the Cleveland sample contains recipients who had not received AFDC since at least January 1990. It is likely that cohorts of new welfare recipients show much greater variation over time than do cohorts of recipients. As economic circumstances change, in particular, new welfare recipients will have greater or less employment history and human capital and, therefore, will have an easier or harder time leaving welfare for employment. In addition, because it is known only that a new recipient had not received welfare since 1990, earlier cohorts in the Cleveland sample may contain some individuals who had received welfare only two years prior, whereas a cohort of entrants in 1994 would contain individuals who had been away from welfare for at least four years. Thus, the earlier cohorts are more likely to contain long-term recipients; therefore, they are likely to have higher subsequent welfare receipt and lower employment.

Figure 10

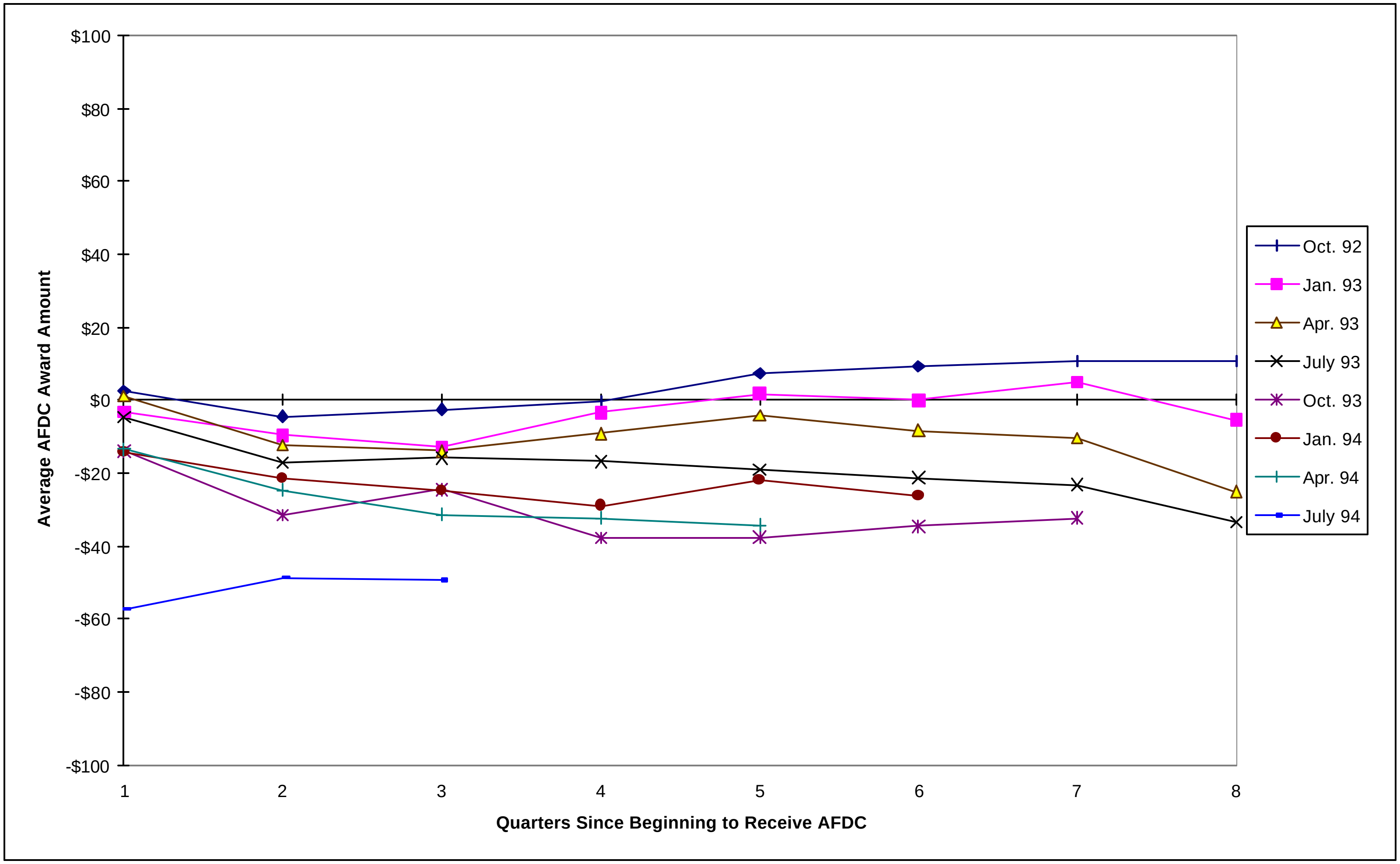
Percentage of Recipients Still Receiving Welfare, by Cohort:
Difference from April 1993 Cohort,
Cleveland AFDC-Entry Cohorts



Source: Cuyahoga County Department of Employment Services, Cleveland, Ohio.

Figure 11

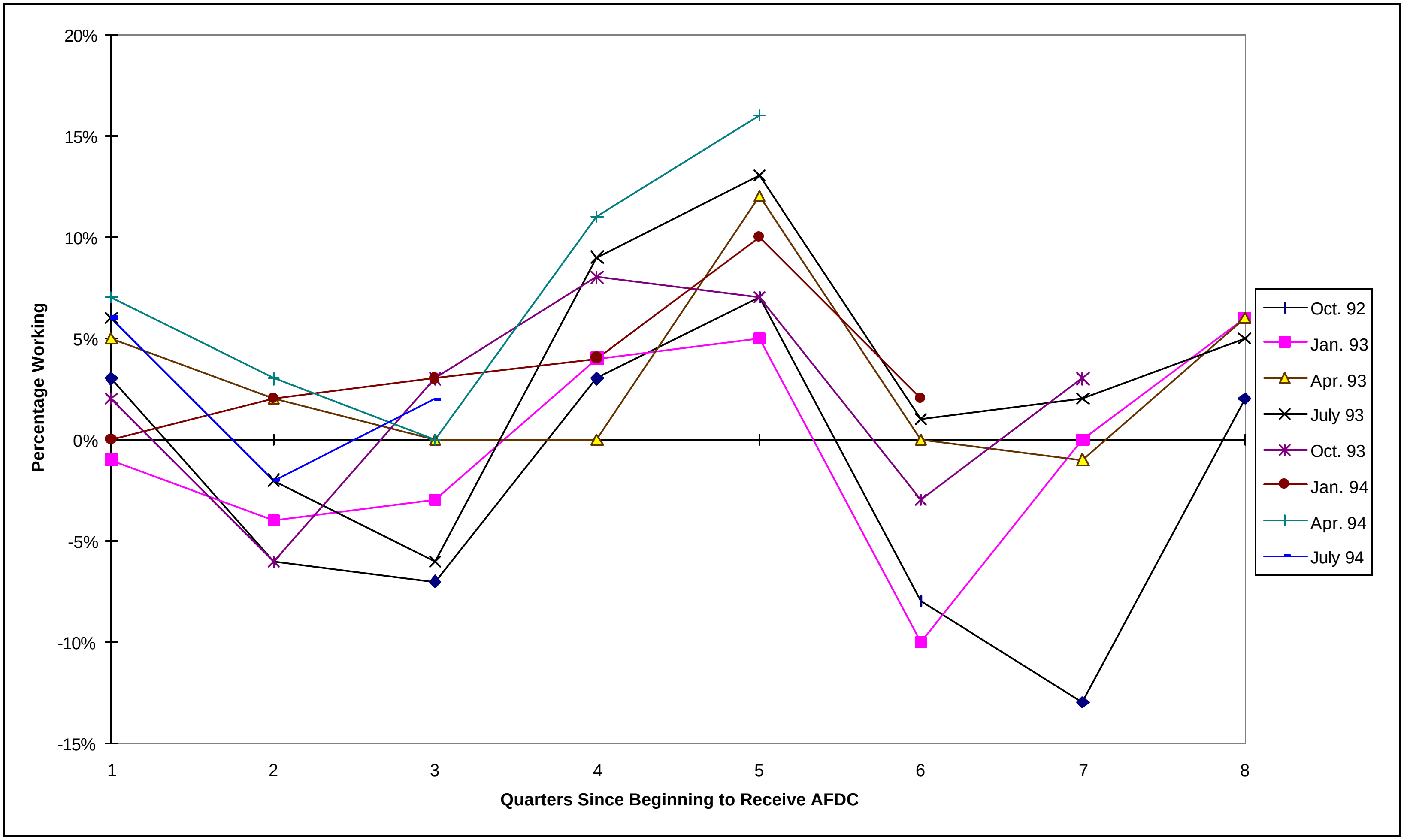
Average Quarterly AFDC Award Amounts, by Cohort:
Difference from April 1993 Cohort,
Cleveland AFDC-Entry Cohorts



NOTE: Average AFDC award includes cases with zero award.
SOURCE: Cuyahoga County Department of Employment Services, Cleveland, Ohio.

Figure 12

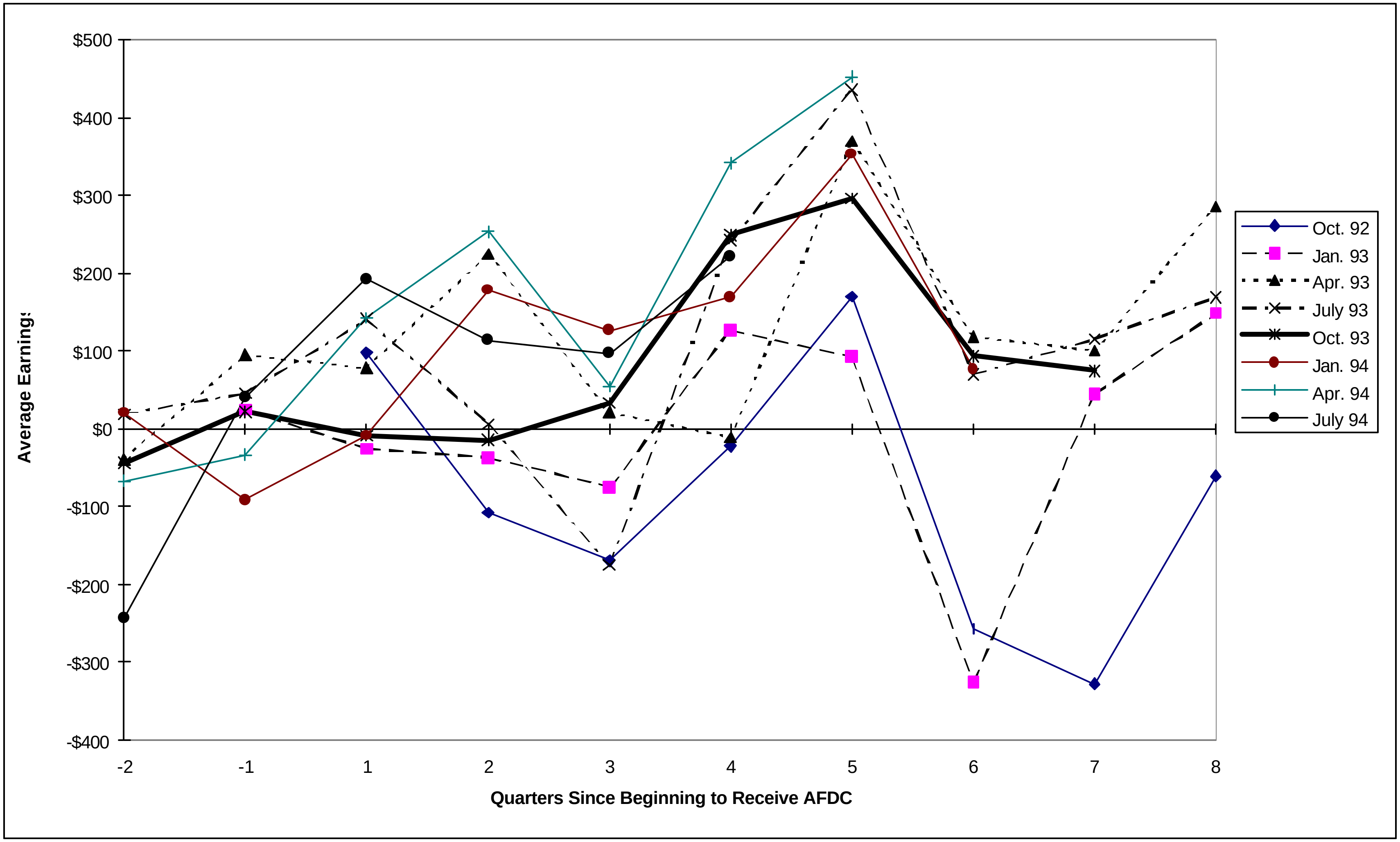
Percentage of Recipients Working, by Cohort:
Difference from April 1993 Cohort,
Cleveland AFDC-Entry Cohorts



SOURCES: Ohio Bureau of Employment Services and Cuyahoga County Department of Employment Services, Cleveland, Ohio.

Figure 13

Average Quarterly Earnings, by Cohort:
Difference from April 1993 Cohort
Cleveland AFDC-Entry Cohorts



NOTE: Average Earnings includes cases with zero earnings.
SOURCES: Ohio Bureau of Employment Services and Cuyahoga County Department of Employment Services, Cleveland, Ohio.

VI. Selecting Cohorts

In an ideal analysis, a cohort would consist of all individuals who would have been eligible for AFDC. Individuals receiving TANF would provide outcomes for those subject to the changes in welfare, while those who would have received AFDC in the absence of reform would compose the counterfactual. Although no analysis can assemble this counterfactual, the Urban Change impact analysis will attempt to define similar cohorts before and after reform by choosing all individuals who received either AFDC/TANF *or* Food Stamps during some period of time. The more of these cohorts that can be assembled, the higher will be the quality of the analysis. Both issues — how the cohorts will be selected and the number of cohorts — introduce potential problems into the analysis.

Sample Selection Bias

Choosing cohorts based on when individuals received welfare introduces problems concerning the comparability of cohorts. If people self-select into a certain population, cohorts drawn from that population will reflect the underlying selection process. To the extent that this selection process is unrelated to the program being evaluated, there is no problem. If, on the other hand, welfare reform affects who goes on welfare, a comparison of outcomes for welfare recipients in different years may suffer from severe selection bias.

Figure 14 graphically illustrates how using Food Stamp data to select the cohorts improves the quality of the analysis. The two identical shaded boxes represent the relevant part of the Food Stamp caseload, which is assumed to be mostly unaffected by the changes in the welfare environment. Inside these two boxes, two circles represent the Food Stamp recipients who are also receiving AFDC (Cohort 1) or TANF (Cohort 2). The smaller size of the second circle represents a hypothesized reduction in caseload because of a deterrence effect associated with the welfare changes. From this figure it is easy to see how a comparison of outcomes for individuals selected from the boxes is a better comparison than one for individuals selected from the circles. As the circles change in size, so will the underlying characteristics of the people within them.

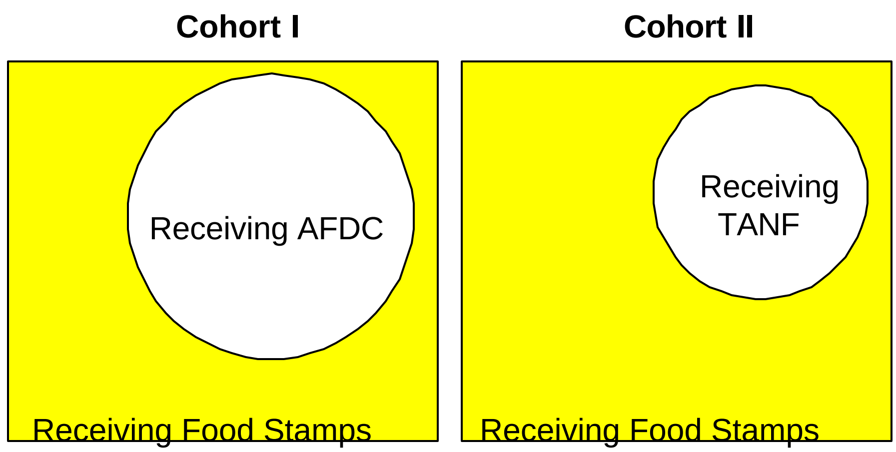
Despite the advantages of using Food Stamp data to do this analysis, it raises two further problems. First, the proportion of nonelderly Food Stamp recipients who are parents and therefore potentially eligible for AFDC, but who are not receiving it, is relatively small.[7] This means that the potential deterrence effects of the new legislation are estimated on a small slice of the potentially eligible population — at least in states with relatively high grant levels (where most Food Stamp recipients with children are also eligible for AFDC, regardless of whether they are working or not). However, in a low-grant state like Florida, the "Food Stamp only" population may be larger, making the Food Stamp caseload a more appropriate sampling frame.

A second problem with using the Food Stamp caseload as a sampling frame is that welfare reform may affect who chooses to apply for Food Stamps; that is, it may affect the size and composition of Cohort 2 in Figure 14. The most substantial change in this regard is the proposed elimination of Food Stamp benefits for immigrants. However, when the sampling frame is defined to include only "potentially AFDC-eligible" Food Stamp recipients, this change may not affect the analysis a great deal (at least

[7]In Figure 14, this is the part of the boxes that is not covered by the circles.

Figure 14

Selecting Cohorts from Food Stamp Data



as far as welfare outcomes are concerned). This problem would become more serious if many people who would have been AFDC recipients in Cohort 1 (that is, included in the first circle) were to seek to avoid contact with the welfare system altogether and would no longer receive Food Stamps either. In that case, these people would disappear from Cohort 2 (the post-reform Food Stamp cohort) and their outcomes would not count in the analysis, causing selection biases similar to those associated with a simple comparison of welfare cohorts. This problem is illustrated in Figure 15.

Aside from trying to use demographic characteristics to make the two boxes (cohorts) look similar, the only way to address this issue with the proposed data is to study changes in the percentage of people in various demographic groups who receive welfare over time.

The Number and Breadth of Cohorts

There are two additional decisions to be made in choosing cohorts: the number of cohorts and the breadth of each cohort. The number of cohorts is determined by how often cohorts are defined, while the breadth is determined by the length of the time period over which the cohort is defined. For example, a cohort might comprise all individuals who received AFDC, TANF, or Food Stamps at any time within a 12-month period. The 12-month period defines the breadth of the cohort. Given this breadth, however, any number of cohorts could be defined. Researchers could choose one such cohort for each year or could increase the number of cohorts twelvefold by choosing one such cohort for each month.

In principle, the greater the number of cohorts, the higher the quality of the estimates of the analysis. More cohorts would yield more precise estimates of both pre-reform and post-reform trends, thereby increasing the ability to tell whether the two trends are different. More cohorts would also improve the ability to estimate the normal variation from period to period and thus determine whether the changes arising from reform are larger than this normal variation. At one extreme, a cohort analysis with just one pre-reform cohort and just one post-reform cohort would neither be able to get an estimate of the trends nor be able to compare changes between cohorts and normal variation.

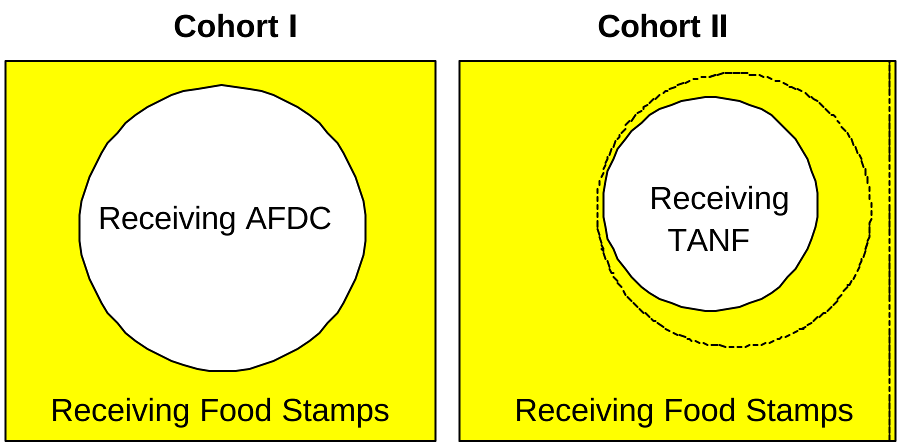
Adding cohorts also artificially introduces stability across cohorts. Consider again the example in which cohorts consist of individuals receiving benefits sometime in a 12-month period, and in which one cohort is defined for each month. The members of one cohort will be identical to the members of the next cohort except for individuals who receive AFDC, TANF, or Food Stamps for only one month in the 12-month period.

Greater turnover across cohorts can be introduced by using a shorter breadth. For example, a cohort could be defined as anyone receiving AFDC, TANF, or Food Stamps within a particular month. If one such cohort is again defined for each month, one cohort will differ from the next for each individual who received AFDC, TANF, or Food Stamps for just one month. Although this would introduce more variation into successive cohorts, it would reduce the number of at-risk individuals included in any cohort and, therefore, would increase potential problems arising from the changing composition of caseloads.

As this discussion indicates, the best definition of a cohort will depend on the percentage of Food Stamps recipients who are not receiving AFDC or TANF, as well as on the rate at which

Figure 15

Compositional Changes in the Food Stamp Caseload



individuals leave welfare. The plan for the impact analysis is to determine the best definition of a cohort only after the administrative data have been examined.

VII. Data Sources

Administrative Records

The analysis of the impacts of welfare reform will rely on administrative data. From the welfare offices at each site will come information about welfare benefits. From the Unemployment Insurance system in each state will come information about earnings and employment.

Welfare participation. To compile historical and prospective records of receipt of welfare payments, administrative records will be obtained for the following programs: AFDC, ADC-U, Food Stamps, and future assistance programs under TANF. Monthly records will be obtained from 1992 through 2002, providing several years of pre-reform welfare history and approximately five years of post-reform tracking for individuals on welfare in 1997. The key data required from these files will include some basic identifiers (such as name, address, date of birth, case number, recipient number, Social Security number) and other identifying information that may be relevant to file linkage. In addition, the file will contain demographic information (such as marital status and ethnic background of case heads) and information on case composition (such as the number of adults and children on the case and the ages of the children).

Ten years of welfare benefit records for every individual provide a rich source of information, but administrative data have an important limitation. Information is available only if the individual lived in the site covered by the administrative agency. For Los Angeles and Cleveland, records will be available for AFDC and Food Stamp benefits paid to county residents. For Miami and Philadelphia, where welfare records are maintained by the state, there is the potential to know whether an individual received welfare and Food Stamps, and how much, as long as the case head lived in Florida and Pennsylvania. Provided that welfare reform does not affect who moves into or away from the four sites under study, this limitation should not affect the ability to discern the direction of reform's effects, although it limits the ability to determine the magnitude of the effects.

Employment and earnings data. Automated Unemployment Insurance (UI) quarterly earnings records will be used to measure workforce attachment, employment spells, duration of employment, and earnings in the formal labor force. For any individual who ever received welfare between 1992 and 2002 in one of the four sites, UI earnings data can be obtained for each quarter from 1992 through 2002.

As administrative data, UI earnings have the same drawback as welfare records. If a person works outside the four states under study, UI records will show zero earnings; so someone who lives or works outside the state will appear to be the same as someone who has stopped working. Again, however, as long as welfare reform does not change who moves, this limitation does not affect the ability to determine whether welfare reform affects earnings and employment.

UI data are further limited because they do not capture the significant amount of work activity that occurs in the informal labor market,[8] which could result in an underestimate of the work activity and earnings of welfare recipients, who have some incentive to hide earnings from the welfare system. This factor is especially problematic in analyzing changes in earnings, because TANF's enhanced disregards may reduce the incentive to hide earnings.

Other Sources of Information

Although administrative data are rich, particularly when millions of cases are being used, they provide limited information. Other components of the Urban Change project will provide key supplemental information for the impact study.

Recipient survey. A second part of the individual impact study is a 90-minute in-person interview that will query about information in a multitude of areas, including education and training participation, labor force and employment, marriage history and household composition, income for each household member from a variety of sources, material hardship, welfare experiences, health and health care, fertility and childbearing, psychological well-being, demographic characteristics, child care, children's outcomes, academic progress of adolescent children, home environment, parenting, absent-father involvement, and child health. In addition, a self-administered questionnaire will gather information from any adolescent children in each family.

In each city, 2,000 mothers will be interviewed for the survey. The first group, containing 1,000 families in each city, will contain mothers who were receiving AFDC or Food Stamps in May 1995, who had minor children at that time, and who were between 20 and 45 years of age at that time. The second group will be identical but will contain mothers who were receiving AFDC or Food Stamps in May 1997. Each group of women will be surveyed once — the first group in early 1998, and the second group in 1999.

The survey also serves several purposes related to the analysis of administrative data. Administrative data will provide accurate information on AFDC, TANF, Food Stamps, and earnings as long as an individual stays in the jurisdiction being studied and as long as earnings are covered by the state Unemployment Insurance system. For the 8,000 mothers interviewed, the survey will provide more detailed information on earnings, public benefits, and other income sources. Welfare reform can have an impact on a family or a neighborhood only if the amount of public assistance dollars going to a family or neighborhood changes. By providing more detailed information about which households lost public benefits and which were able to replace public benefits, the survey will more precisely link benefits to employment and other outcomes. Administrative data will indicate whether a person stopped receiving welfare, but not why. The survey will provide richer information on the recipients' perspectives and whether benefits were lost due to sanction or whether a recipient chose to leave welfare to avoid increased obligations.

[8]In interviewing several hundred welfare recipients in Boston, Chicago, San Antonio, and Charleston, Edin and Lein (1997) found that virtually all had additional sources of income intentionally hidden from caseworkers.

Ethnographic study. Although the recipient survey will fill in some gaps, the Urban Change project also contains an ethnographic component to provide even more detailed information about a small group of families — 10 families from three neighborhoods in each of the four cities, for a total of 120 families. For this small number of families, the ethnographic study will gather a broad range of information about their experiences in going through welfare reform, their methods of coping with what is presumably a less generous public assistance environment, and their outcomes under the new policy. The ethnographic study will assist in interpreting results of the impact analysis by providing more accurate and more detailed information about recipients' earnings, income sources, and expenditures.

Neighborhood indicators study. This component of the Urban Change project will provide valuable information to answer questions about the surrounding communities. One of the drawbacks of administrative data is that they will reveal nothing about families who receive neither welfare nor Food Stamps between 1992 and 2002. By combining estimates of the population of each census tract with the number of individuals receiving welfare, the neighborhood indicators study will help track the *percentage* of the population on welfare in an area. While administrative data will disclose whether welfare recipients are working more or earning more, the neighborhood indicators component will describe the economic vitality of neighborhoods with high concentrations of welfare recipients. Finally, by following crime rates and other measures of social well-being, the neighborhood indicators study will probe the effects of welfare reform on wider social measures.

Implementation study. This component of the Urban Change project will yield crucial information about the timing of reform. Although legislation indicates when reform is supposed to begin, local administrators have some control over how that reform is implemented. The better one can pinpoint the timing of reform and the strictness of time limits — whether and when the local office used informal roadblocks to discourage welfare receipt, whether the local office tried to convince recipients that they should work — the more likely that reform's effects can be disentangled from unrelated changes.

VIII. Conclusions

One objective of MDRC's Urban Change project, which is studying the devolution of welfare in four cities, is to compare what occurs under devolution with estimates of what would have occurred under the superseded rules of AFDC. Using the universe of individuals who ever received Food Stamps, AFDC, or TANF from 1992 through 2002, the individual impact study will assign individuals to cohorts of welfare recipients. By defining and following many of these cohorts over time, the nonexperimental analysis offers many of the advantages of time-series analyses, particularly the ability to adjust for complicated pre-reform trends and the ability to determine how much variation from time to time is normal. By following cohorts of individuals, the analysis also offers many of the advantages of studies that compare several cross-sections over time, namely, the ability to adjust for demographic changes in the population and to correct for maturation.

Although the multiple cohort design is the best alternative for analyzing the impacts of devolution, it is not foolproof. For example, using data from MDRC's evaluation of Project Independence — Florida's JOBS program — a multiple cohort analysis produces estimates of the impacts on earnings and AFDC benefits that are off by more than $100 in some quarters, and impacts on employment rates and AFDC recipiency rates that are off by 3 to 4 percentage points in some quarters.

To probe further for bias, two other sources of data were used: individuals from an evaluation of the JTPA and cohorts of new welfare recipients in Cleveland. Across JTPA cohorts, monthly earnings differ by as much as $40 per month, or $120 for a quarter of such months, while employment rates differ by as much as 5 percentage points per month. Across the Cleveland cohorts, variation in earnings is more than $400 in some quarters, while variation in employment and AFDC recipiency rates are nearly 15 percentage points in some quarters.

Together, these three sets of results imply that the natural variation over time might be as high as $400 per quarter; the natural variation in employment rates, as high as 15 percentage points; the natural variation in AFDC benefits, as high as $100; and the natural variation in AFDC recipiency rates, as high as 15 percentage points. Thus, if the Urban Change impact analysis finds lesser effects, one could not be sure that they resulted from devolution and not from random chance.

Appendix A

Econometrics of the Multiple Cohort Design

This appendix provides more technical details about how the multiple cohort design for the Urban Change impact analysis could be implemented using regression analysis. These details should not be viewed as necessarily indicating how the final analysis will be conducted, but they represent a starting point.

In this impact analysis, a cohort will contain a group of individuals who received AFDC/TANF or Food Stamps in some period of time, probably a year.[9] Two types of evidence could imply that TANF has affected people's behavior. First, cohorts observed prior to TANF may have different outcomes than similar cohorts observed after TANF policies have been implemented. Second, all cohorts may markedly change their behavior around the time that the entire TANF policy or specific policies such as time limits are implemented.

The data sets proposed for this analysis will be enormous — including information on several million people in Los Angeles County, for example. Such large data sets will yield extremely precise estimates, even of effects that are relatively small. This appendix also addresses a potential method for asking whether the estimated effects of TANF are small. To do this, subcohorts would be defined based on demographic factors, particularly age and race. It is expected that similar subcohorts will be affected about the same by TANF, and that very different subcohorts will be affected more differently. For example, TANF should have similar effects for African-American women ages 20-22 as for African-American women ages 23-25, but very different effects for white women ages 35-40. If, in contrast to this expectation, quite different effects are found for similar cohorts, that will be an indication that the estimated effects of TANF are not reliable estimates of the true effects of the policy change.

The Model

In the Urban Change impact analysis, a cohort will be defined by two features: when its members were chosen and some demographic characteristic or characteristics such as age or race. For example, one cohort might contain sample members who were receiving AFDC or Food Stamps in 1992 and who were also ages 20-22 in 1992. Or a cohort might consist of all African-Americans receiving welfare in 1994. There will be many such cohorts, perhaps one for each year and for each demographic subgroup, depending on how many individuals are in the sample.

In the notation that follows, let j refer to the time period during which the cohort was chosen (1992 and 1994 in the examples above) and z refer to the characteristic shared by everyone in the cohort (ages 20-22 and African-Americans in the examples). In addition, let the subscript i indicate a measure for a particular individual and the subscript t indicate a particular calendar period. For example, the notation y_{ijzt} indicates the outcome for individual i in year t who was taken from a cohort chosen at time j that included only individuals with characteristic z. The difference between when an outcome is

[9]In some sites, the sample might also include people who were eligible for Medicaid, even if they did not receive AFDC, TANF, or Food Stamps. This is most likely to be the case in Cuyahoga County and Los Angeles County.

being measured (t) and when the cohort was defined (j) will be called the *maturity* of the cohort at time t and is equal to t-j. Using this notation, an individual's outcome could be written as shown below:

$$y_{ijzt} = \gamma_{t-j,z} + f_i + TANF_t\,\eta_{t-j,z} + \delta_t + \beta X_{it} + \varepsilon_{ijzt}$$

In this equation,

- y_{ijzt} is an outcome at time t for person i, who is in a cohort chosen at time j with characteristic z.
- $\gamma_{j-t,z}$ is an intercept that applies to people from a cohort with characteristic z of maturity t-j at time t.
- f_i is a person-specific effect.
- $TANF_t$ is a binary variable equal to 1 for time periods after TANF was implemented, with $\eta_{t-j,z}$ being the corresponding effect of TANF on a cohort of maturity t-j with demographic factor z.
- δ_t represents all unobserved factors that affect outcomes for all cohorts at time t.
- t is a time trend that represents the amount of time that has elapsed since some fixed time, such as the beginning of 1992.
- X_{it} is a vector of individual characteristics which vary over time, with β being the corresponding effect of those demographics.
- ε_{ijzt} is a random error that is independent and identically distributed (i.i.d.) across individuals, cohorts, and time.

In the Urban Change impact study, the outcome of interest (the y) might include whether the individual received AFDC/TANF in a month or quarter, whether she received earnings in a job covered by the state's Unemployment Insurance system, and what amounts of welfare benefits and income were received. The administrative records used for the study will contain limited demographic information. However, X_{it} could include such factors as age — unless age is used to define the cohort (that is, unless age is part of z). Because the equation includes a person-specific fixed effect, it is impossible at this stage to determine the effect on outcomes of characteristics that do not change over time.

This is a somewhat complicated equation that can be understood best by recognizing what the model should be accomplishing:

- **Removing the effects of demographic differences across cohorts.** A cohort comparison works best when the cohorts being compared are similar, so that differences in outcomes can be attributed more reasonably to differences in policy rather than to differences in the demographic characteristics of the cohorts. There is no reason to expect cohorts in the Urban Change project to be similar, particularly since the study covers such a long period of time. The American population has been aging, and that might be reflected in differences in average age over time. The economy has generally improved in all four Urban Change sites since 1992, and that presumably affects who enters welfare at different points in time. One purpose of the model, therefore, is to remove demographic differences before the next steps look for the effects of TANF. This task is accomplished in three parts. First, including the time-varying covariates X_{it} controls for some factors such as age and number of children. Second, allowing for person-specific effects removes the effects of unobserved demographic differences across individuals and cohorts. Third, defining a cohort based on demographics z means that comparisons can be made among successive cohorts with similar demographics.

- **Removing the effects of economic conditions.** In the ongoing debate about why welfare caseloads have declined since 1993, the two leading explanations that have been offered are welfare reform and economic conditions. To separate the effects of TANF from the effects of the economy, the model must include some measures of economic conditions, such as unemployment rates or indicators of job growth. This is accomplished in the model by including the economic factors represented by u_t.

- **Determining whether patterns of outcomes for cohorts after welfare reform are different from patterns for earlier cohorts observed before welfare reform.** Suppose that two cohorts are compared in Miami-Dade County. One was first observed in 1992, while the other was first observed in 1996, when Florida's welfare reform was implemented. For the 1992 cohort, welfare receipt will probably be somewhat lower in 1993 than in 1992. Likewise, for the 1996 cohort, welfare receipt will probably be somewhat lower in 1997 than in 1996. Will the decline for the 1996 cohort be greater than the decline for the 1992 cohort? More generally, will the first-year declines in welfare receipt for cohorts defined after welfare reform be greater than the first-year declines for welfare receipt for cohorts defined before welfare reform? What about the second-year changes in welfare receipt, and the third-year changes, and so on? The model accomplishes this task through the parameters represented by $\eta_{t-j,z}$. If TANF is having an effect, then these coefficients jointly should be significantly different from zero.

- **Removing the effects of ongoing time trends that predate welfare reform.** In the comparison of the 1996 and 1992 cohorts described above, suppose that the first-year decline in welfare receipt for the 1996 cohort is greater than for the 1992 cohort. One interpretation is that TANF had an effect — getting people to leave welfare faster than they had before TANF. A second interpretation is that the steeper decline in welfare use for the 1996 cohort reflected trends that were ongoing before the TANF reform and that helped explain the decline in caseloads since 1993. To rule out these effects, the model must adjust for pre-TANF trends before looking for the effects of TANF. This task is accomplished in the model by including the time dummy variables (with estimates indicated by δ_t).

Estimating the Model with Overlapping Cohorts

One question that arises in estimating this model is how to handle the issue of overlapping cohorts, that is, cohorts which are defined so that one individual can be part of more than one cohort. One cohort might include anyone who received AFDC or Food Stamps in 1992, and a second cohort might include anyone who received AFDC or Food Stamps in 1993. The two cohorts will contain many of the same people, because many recipients of public assistance receive benefits for more than one year. Although overlapping cohorts might seem difficult to handle, the econometrics of overlapping cohorts is best thought of as an extension of the econometrics of non-overlapping cohorts. When cohorts do not overlap, an individual is still observed during more than one time period, and the fact that an individual contributes to knowledge at different times must be accounted for. In other words, the model implies that unobserved factors in one time period for a cohort will be correlated with unobserved factors in other time periods for the same cohort. Likewise, when cohorts overlap, an individual contributes not only to estimates that refer to different time periods but also to estimates that refer to different cohorts.[10] In this case, the unobserved factors for one cohort will therefore be correlated with the unobserved factors for other cohorts, because they will include some of the same individuals. Stated in this way, however, it is clear that the estimation of the model with overlapping cohorts is simply a type of very complicated generalized least squares (GLS) regression model in which all error terms are correlated with all other terms.

[10]This ignores the obvious degenerate case in which one cohort contains exactly the same individuals as a second cohort and assumes that cohorts are defined so that the overlap is not complete. In other words, it assumes that some individuals in a cohort will be in only that cohort. This will be true for all individuals who receive public assistance in only one calendar year.

References

Cook, Thomas, and Donald Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand-McNally.

Edin, Kathryn, and Laura Lein. 1997. *Making Ends Meet.* New York: Russell Sage Foundation.

Friedlander, Daniel, and Philip K. Robins. 1995. "Evaluating Program Evaluations: New Evidence on Commonly Used Nonexperimental Methods." *American Economic Review* 85(4).

Kemple, James, Daniel Friedlander, and Veronica Fellerath. 1995. *Florida's Project Independence: Benefits, Costs, and Two-Year Impacts of Florida's JOBS Program.* New York: Manpower Demonstration Research Corporation.