Conference Series No. 30

Lessons from the Income Maintenance Experiments

Proceedings of a Conference Held in September 1986

Alicia H. Munnell, Editor

Sponsored by:

Federal Reserve Bank of Boston and The Brookings Institution

Ashenfelter
Blum
Bradbury
Burtless
Cain
Cohen
Coyle
Ellwood
Elmore
Gramlich
Hall
Hanushek
Hausman
Heclo
Mead
Metcalf
Michael
Murray
Nathan
Rainwater
Reischauer
Rossi
Solow
Tuma
Wildavsky
Zellner

Lessons from the Income Maintenance Experiments

Proceedings of a Conference
Held at Melvin Village, New Hampshire
September 1986

Alicia H. Munnell, Editor

Sponsored by:

Federal Reserve Bank of Boston
and
The Brookings Institution

Contents

The Methodology

The Historical and Policy Context

The Policy Lessons

Lessons from the Income Maintenance Experiments: An Overview

*Alicia H. Munnell**

The United States public welfare system has been a source of discontent for many years. The system has been characterized as one that discourages work, undermines the family, and perpetuates dependence. In the late 1960s and early 1970s, many experts believed that the negative income tax represented a simple and desirable alternative to the existing programs. The complex set of cash and in-kind benefits paid to certain categories of the poor would be replaced with a single guaranteed income payment for all poor families that would gradually diminish as earnings increased.

Congress, however, was extremely reluctant to enact such a plan. One reason for the political opposition was the widespread fear that a guaranteed income would reduce the work effort of poor breadwinners and, as a result, cost taxpayers a great deal of money. In an effort to gain some knowledge about the potential impact of a guaranteed income on labor force activity, the federal government in the late 1960s and 1970s sponsored four large-scale social experiments to measure individuals' responses to different levels of benefits and tax rates. Although the negative income tax itself has fallen from favor, the labor supply question and the other basic issues studied in these experiments are still relevant to the current social welfare debates. Architects of new programs need to know the effects of particular reforms on work effort, family stability, housing, food consumption and the well-being of dependent children.

The negative income tax was tested in four separate experiments. The first experiment, in New Jersey and Pennsylvania, lasted from 1968

*Senior Vice President and Director of Research, Federal Reserve Bank of Boston.

until 1972 and had a sample size of 1,357 households, consisting of low-income couples from declining urban areas. The rural experiment, which was conducted in Iowa and North Carolina from 1969 to 1973, included 809 low-income rural families. The third experiment, which took place in Gary, Indiana between 1971 and 1974, was composed of 1,780 black households, 59 percent of which were headed by single females. The largest experiment, which contained 4,800 families, was conducted in Seattle and Denver from 1971 to 1982. The Seattle-Denver experiment not only offered recipients more generous plans than the other experiments, but also extended the duration from three to five years for a quarter of the participants.

Although the last of the four experiments ended in 1982, the major lessons of the experiments are neither widely known nor well understood. Indeed, the final reports from the two largest and most important experiments—those in Gary, Indiana and in Seattle and Denver—have never been published in a broadly accessible form. The experiments also represent a landmark in the history of social policy. The New Jersey experiment was the first large-scale attempt to test a policy initiative by randomly assigning individuals to alternative programs, and random assignment of participants to treatment and to control groups was an important feature of all four experiments. The procedure reduces the possibility of bias toward the tested plan on the part of sponsors and researchers. Although some of the results of the experiments are not conclusive and are the subject of vigorous debate among specialists, the experience gained from the undertaking offers valuable lessons for future policy research projects. Both to summarize the findings and to derive the methodological and policy lessons, the Federal Reserve Bank of Boston and The Brookings Institution jointly sponsored, in the fall of 1986, a conference on "Lessons from the Income Maintenance Experiments," the results of which are published in this volume.

The first set of three papers reexamines the empirical findings on labor supply response, family stability and a host of other factors, such as consumption, investment, and child well-being. While most of the reworking of the data yields results similar to those previously published, no consistent and reliable support is found for the earlier indications of large increases in the family breakup rate for those eligible for guaranteed income payments in the Seattle-Denver experiment. This new result is very important, since the threat of family dissolution is frequently used as an argument against guaranteed income payments.

The empirical papers are followed by a critical assessment of the methodology of the social experiments and the credibility of the main findings. The experiments are then placed in historical context to examine why and how they came into existence and their contribution to the policy debates. Following this analysis is a series of papers on policy lessons from the experiments as viewed from the perspectives of a sociologist, a political scientist, an economist, and a public administrator. A concluding paper summarizes the implications of these lessons for future efforts to reform the welfare system.

What Do the Experiments Tell Us?

Data from the four negative income tax experiments were used to analyze the effects of various combinations of guaranteed payments and tax rates on labor supply, family stability and a host of peripheral issues. The following papers show that the results for labor supply responses are quite robust across sites, populations, and treatments, whereas the widely publicized conclusions on marital stability fail to hold up under closer scrutiny. Although the experiments were not designed to yield high-quality data on consumption patterns and other factors, the suggestive results for these peripheral effects provide useful insights.

Labor Supply

Gary Burtless reported two different types of labor supply estimates. The first was the simple difference between the work effort of people who were assigned to the experimental programs and those who were assigned to the control groups. Generally, the experiments caused moderate reductions in work effort. The responses were greater among women (an average reduction of 17 percent) than among men (7 percent). The largest absolute reductions occurred in the Seattle-Denver experiment, which offered the most generous plans. These work effort responses were overstated to the extent that participants underreported their earnings in order to receive larger benefits, but understated to the extent that a limited duration experiment elicits a smaller response than would be expected from an equivalent permanent program. This was particularly the case for plans with high guaranteed incomes and low tax rates.

Because estimates of average responses in specific experiments are difficult to use for predicting the consequences of alternative national reform proposals, Burtless also reported structural estimates of response. Weighted averages of income and substitution elasticities from the four experiments imply a much smaller responsiveness to guaranteed income disincentives than do most nonexperimental estimates, and they also fall in a far narrower range.

Burtless concluded by presenting the results of microsimulations using elasticity estimates from the Seattle-Denver experiment to calculate work effort response and budgetary costs for the nation as a whole under alternative negative income tax plans. The results highlight a conflict between the goal of providing work incentives to transfer recipients and that of providing incentives to the population as a whole. Recipients can be encouraged to work by reducing the tax rate applied to benefits as earnings rise, but such a reduction will increase the number of benefit recipients and hence reduce aggregate work incentives. In terms of budgetary implications, a plan that offers guarantees equal to the poverty line with a moderate tax rate would cost roughly $60 billion more than current welfare and food stamp programs; this figure falls to roughly $20 billion with a higher tax rate.

While it appears that poverty could be eliminated at relatively modest cost under the less ambitious plan, the labor supply responses indicate that earnings reductions would offset at least part of the income gains to the poor produced by the plan. As much as 40 to 58 percent of the added transfers for two-parent families would be offset by earnings reductions on the part of husbands and wives. The problem is less severe in the case of single mothers, where earnings would fall by only 16 to 20 percent of additional costs.

In short, the four income maintenance experiments showed that guaranteed incomes reduced work effort. The reductions were probably larger than advocates had hoped, but considerably smaller and more precisely measured than predictions based on prior nonexperimental research. Even though the overall work reduction is small, the resulting earnings loss among recipient breadwinners would represent a large fraction of the payments to low-income families. This is a significant political impediment to trying to reduce poverty through a system of pure cash transfers.

Burtless's formal discussants raised some serious concerns about his assessment of the labor supply responses. Orley Ashenfelter's first point pertained to Burtless's conclusion that a reduction in work effort due to underreporting is just as costly to taxpayers as a genuine reduction in work effort. Ashenfelter contended that an equally plausible conclusion is that a real nationwide negative income tax would operate using government reports on income and therefore would involve little cost from underreporting. The real problem in his view was that the experiments were not designed to address the possibility of underreporting, so it is impossible to tell from the data whether a genuine scheme would produce a labor supply response, further underreporting, or neither.

Ashenfelter's second point related to estimating the magnitudes of the income and the substitution effects; the experiments provided no direct information on the question of whether higher tax rates led to greater labor supply response or whether more generous payments induced a larger reduction in work effort. Instead, the values for the income and substitution effects were delivered from models that Ashenfelter feared primarily reflected the prior beliefs of the investigators.

Robert Hall made three points. First, in those cases where nonexperimental data from the unemployment insurance system confirm substantial underreporting, the labor supply responses should be studied directly using those data. Second, the smaller substitution and income effects in the experimental studies tend to confirm that the results of nonexperimental studies are tainted by the high correlation between wages and preferences for working. Finally, Hall criticized Burtless's evaluation of negative income tax programs in terms of the ratio of earnings reductions to program "costs."

Family Stability

Glen Cain reexamined the evidence from the experiments on the issue of family stability. He concentrated on a 1983 study conducted by Groeneveld, Hannan, and Tuma, which had produced the startling finding that the negative income tax dramatically increased marital dissolutions.

In theory, according to Cain, a negative income tax that was equally as generous as the existing welfare program—namely, aid to families with dependent children (AFDC)—would be expected to promote marital stability. The negative income tax would provide the same benefit as AFDC to a separated or divorced mother and more than AFDC to a married woman and her husband, so that it would reduce the price subsidy to divorce. Moreover, because the negative income tax provides benefits to intact families, while AFDC frequently does not, it produces higher family incomes, which are presumed to have a positive impact on marital stability. A negative income tax that is less generous than the AFDC program still reduces the price subsidy to divorce and has a pro-stability income effect, albeit smaller. In the case of a negative income tax plan that is more generous than the existing AFDC program, the predicted effects are ambiguous. The pure income effect promotes marital stability, while the net price effect would probably encourage divorce. (Although the payment for both the divorced woman and for the woman and her husband would be higher under the more generous plan than under AFDC, the higher level of payments to the woman is presumed to dominate the comparisons in her decision to remain married or to become divorced.)

Groeneveld, Hannan, and Tuma found that the negative income tax plans tested in the Seattle-Denver experiment increased the rate at which marriages dissolved among white and black couples by 40 to 60 percent. One explanation for these results could have been that the relative generosity of the payments in the Seattle-Denver experiment produced negative price effects that dominated the positive income effects. However, this apparently was not the case because the least generous plans, which offered about the same payments as AFDC or lower ones, induced the largest destabilizing effects, while the most generous plans had no adverse impact on marital stability.

Using Groeneveld, Hannan, and Tuma's model and data, Cain was able to duplicate their dramatic results. He then made several modifications to the analysis: he eliminated couples without children (since they would presumably be excluded from any program passed by Congress); he separated the group who received only a negative income tax payment from those who received both the payment and training; and he included information on marital dissolutions even if they occurred after the couple left the experiment. The greatest difference between Cain's analysis and the earlier work, however, was that he included the full five years of the five-year experiment, while Groeneveld, Hannan, and Tuma emphasized results from the first three years.

With these modifications and timing differences, Cain found only small and inconsistent effects on marital stability. In the case of white and Hispanic couples, neither the benefits nor the training nor the interaction of the two had a statistically significant effect on the rate at which marriages were dissolved. For blacks, on the other hand, the impact of the combination of the negative income tax and the training program was destabilizing and statistically significant. In terms of the impact of the pure negative income tax plans (that is, payment without requiring training) on all the groups, half the coefficients indicated a stabilizing effect and half a destabilizing effect, with only one of the coefficients statistically significant. Even when the site and duration samples were aggregated, the only significant effect was the destabilizing impact of the combined benefit and training program on blacks. This led Cain to conclude that "the evidence [about the impact of the negative income tax on marital stability] is not decisive or even persuasive." In any case, Cain argued, short-duration experiments cannot be expected to yield decisive results on demographic behavior, since they do not simulate the incentives of a permanent negative income tax.

In response, Nancy Tuma, one of the authors of the original study, argued that the evidence, while not decisive, was persuasive. Tuma viewed Cain's estimated increase in the marital breakup rate from the pure negative income tax of 17 percent for whites and 31 percent for blacks as large enough to be noteworthy. The lack of statistical significance of the coefficients was to be expected, she argued, in view of the small sample size.

Moreover, she questioned some of Cain's analytical decisions that reduced the negative income tax effects. For example, Tuma acknowledged that the presence of children reduced the response to the negative income tax, but argued that social scientists had a responsibility to analyze all the data. Second, separating the pure negative income tax from the combined benefit and training program reduces the sample size so much that chance variations can swamp major trends. Finally, Cain failed to mention the analysis of pooled data from the Seattle-Denver and New Jersey experiments, which showed statistically significant increases in the rate of marital breakup.

David Ellwood basically agreed with Cain that very little has been learned from the negative income tax experiments about separation and divorce. The evidence indicates that the programs probably were not stabilizing and may have been somewhat destabilizing. This, however, was to be expected given the generosity of negative income tax payments relative to those provided under AFDC. The small sizes in the Seattle-Denver experiment for groupings by race or site or treatment preclude any definitive findings with nationwide application.

Other Effects

Eric Hanushek summarized the impact of negative income tax payments on consumption and investment — specifically, on housing and education choices made by participants in the experiments. He limited his review to these two areas because the experiments were not designed to provide information on non-labor-supply responses and these topics were ones where common findings could be generalized from the four experiments.

A major motive for examining the consumption response is the suspicion by some that the increased income would be spent on frivolous or immoral products, such as fancy cars, color TVs or drugs. On this score, the results should be very comforting to those concerned that the money would be "squandered." Consumption rose modestly, as would be expected with a slight rise in income, but the pattern of expenditures remained unchanged from that which existed in the absence of the payments.

One component of consumption where increases would have been viewed as unambiguously good is housing, but the payments appear to have had little effect on housing expenditures. Instead, the income maintenance experiments (in conjunction with results from the housing allowance experiments) demonstrated that, contrary to the commonly held belief that the income elasticity of housing was approximately one, the elasticities for the poor were quite low: a 10 percent increase in permanent income would lead to an increase in housing expenditures of 2 to 3 percent in the short run and 5 percent in the long run. Results from the Gary and Seattle-Denver experiments did suggest that the income maintenance programs encouraged homeownership, but this result, given the temporary nature of the program, probably reflected a shift in the timing of already planned house purchases.

The most likely place that income maintenance payments would affect investment is the area of human capital, and, with regard to this, analysts have focused on both school attendance and scholastic performance. Although the evidence on scholastic peformance is mixed and weak, the experiments do appear to have affected attendance. A negative income tax would influence the school-attendance decision by reducing the cost of not being in the labor force, and the data from the experiments show that, for the experimental period, the programs did appear to induce more schooling. In fact, the reduction in labor force activity for young people brought about by the negative income tax is almost completely offset by increased school attendance. Hence, the encouragement of skill development may be one of the positive side benefits from the introduction of a negative income tax.

Katharine Bradbury expanded on Hanushek's paper by summarizing the research relating to some other areas of consumption and investment, including health, and social and psychological well-being. She emphasized that findings about how people spend additional income are important not only because they provide some facts to help displace old stereotypes, but also because they can assist policymakers who must choose between cash assistance and targeted forms of aid. For example, as far as the researchers could determine, medical care utilization did not increase and health status did not improve as a result of the income maintenance payments. Hence, to the extent that improved health is of particular interest, programs aimed directly at health care have a better chance of success than do cash transfers. In terms of psychological well-being and participation in community life, again the researchers found no effect. Overall, the results suggest that the lives of recipients were not altered dramatically by the payments offered in the experiments.

Robert Michael reiterated the point that the experiments were ill-suited to yield high-quality data on topics other than labor supply, but argued, nevertheless, that important suggestive results should not be overlooked in any review. For example, studies of the Seattle-Denver experiments showed a substitution toward market forms of child care from family care and other nonmarket forms. The Seattle-Denver experiments also made it possible to study migration, since they permitted recipients who moved to continue receiving benefits; the results showed that the rate of migration was 50 percent higher for those in the experimental negative income tax plans than for the controls. Investigators also looked at the effects of the experiments on fertility using the Seattle-Denver data; in this case, the results were inconclusive since the effect was negative for whites, positive for Hispanics, and not statistically significant for blacks. Michael concluded, however, that while the peripheral results are interesting and provocative, the weakness of the experimental data for investigating these issues has forced researchers to look to alternative data sources for subsequent analysis.

In summary, the survey of empirical findings suggests that the income maintenance experiments caused a moderate but manageable reduction in labor force activity, had no statistically significant stabilizing or destabilizing effect on the marriages of couples with children, and basically did not alter noticeably the consumption and investment decisions of recipients. The question that remains is: how much weight can be placed on these results?

How Reliable Are the Results?

Arnold Zellner and Peter Rossi touched off a heated debate with their sharp criticism of the goals, design, execution, and analysis of the income maintenance experiments. In their opinion, inadequate attention was devoted to formulating clear-cut objectives. For example, to the extent that the goal was to estimate the cost of alternative negative income tax plans, the experiments were not really designed to provide the appropriate information. Feasibility studies or pilot projects were generally nonexistent. Serious measurement problems were not adequately resolved. Design statisticians, survey experts, and other specialists did not play an active enough role in the planning and execution of the experiments. Management and administration procedures were not completely satisfactory. Policymakers and researchers did not share clearly stated objectives. The experimental designs and the models on which they were based were frequently inadequate. Finally, the quality of reporting of results left much to be desired.

The authors made several suggestions for improving the methodology of future experiments. To provide useful predictions, such experiments should employ a sufficiently large national probability sample and test a wider range of treatments. (In the Seattle-Denver experiment, for example, marginal tax rates varying only between 0.5 and 0.8 were employed.) Second, if researchers are uncertain about which model to use, experiments should be designed to provide information to discriminate among the alternatives. Third, randomization should be used, since it mitigates the effects of model misspecification and produces robust statistical designs. Fourth, in view of the considerable uncertainty over how the models should be specified, it is important to test the predictive ability of the models used in the experiments. For example, the labor supply equations from the Seattle experiment could have been used to predict labor response in Denver. Fifth, the results should not be presented simply as point estimates, but rather reported in terms of the probability that the estimates lie within a certain range. Moreover, it is useful to note that if the outcomes for individual experimental units are not independent, the precision of the estimates disappears rapidly. Sixth, recognizing the dynamic aspects of economic behavior leads one to construct models different from the static ones used in the income maintenance experiments; the experiments are of short duration while the policies are permanent and may therefore call forth a different response. Finally, whenever it is feasible, social experiments should be linked to ongoing longitudinal surveys.

Jerry Hausman, the first formal discussant, stressed the authors' point that experiments should provide usable predictions of the effects of various proposed policies and measures of predictive precision. This consideration has two corollaries: First, the experiment should cover the entire range of possible options so that policymakers do not have to extrapolate results to untested plans. Second, the design of the experiment must supply results that are sufficiently precise to be useful. Hausman's greatest disappointment with the results from the negative income tax experiments was the lack of precision. In terms of reporting the results, however, Hausman did not think it was necessary to adopt the Bayesian approach, since he had found that point estimates and standard errors were sufficient for most audiences. Finally, he supported the Zellner-Rossi call for panel data, but noted that the necessity of keeping track of panel members may raise the costs considerably. Overall, Hausman agreed with the Zellner-Rossi conclusion that the goal, design, execution, and analysis of the income maintenance experiments left much to be desired. He attributed the failings, however, to the fact that the Gary and Seattle-Denver experiments were designed and executed before the lessons of the New Jersey experiment were learned.

Charles Metcalf found Zellner and Rossi's recommendations and criticisms naive. For example, their call for interaction between sponsors and bidders in preparing proposals reflects a simplistic view of the competitive procurement process; often the design and execution phases of an experiment are carried out by different organizations under separate contracts. Moreover, a pilot project may not be needed in an environment cluttered with an extensive history of social experiments, especially since pilots may delay the experiment for a considerable period. Additionally, the Zellner-Rossi suggestion that a national sample is absolutely necessary to make national cost estimates fails to recognize the tradeoff often required between the sample being from the relevant population and the intervention tested being relevant in terms of program, duration, and other features. The increasingly prevalent view is that experiments work only if the intervention is carried out by "real" program agencies rather than by experimenters, and this tends to limit the number of jurisdictions that can be covered by an experiment. Finally, Metcalf noted that evidence is mounting that efforts to use longitudinal panels as comparison group alternatives to randomized control groups have been unsuccessful, and rejected Zellner and Rossi's proposal that a longitudinal panel could be used as the basis for drawing experimental samples.

Metcalf also thought that Zellner and Rossi were unrealistic in some of their criticisms. For example, they argued that the experiments should have tested a broader range of plans, a suggestion with which most experimenters would agree from a pure design perspective; but the policymakers financing the experiments were reluctant to consider "extreme" plans outside the "relevant policy range." Zellner and Rossi characterized as "unusual" the use of the status quo rather than "no treatment" as controls, the basis of comparison in social experiments; however, removing the individuals who form the control group from AFDC would be an extremely unrealistic definition of no treatment. Moreover, one of the objectives of the study was to provide internally valid direct estimates of the relative costs of AFDC and the negative income tax. Finally, Metcalf argued that Zellner and Rossi's effort to discredit the nominal standard errors from the experiment by alluding to cross-unit dependence was extremely misleading.

The discussion of the Zellner-Rossi paper was heated. Robert Spiegelman called many of the authors' direct and implied criticisms "off base." He argued that the experiments did have a clearly defined objective — namely, to measure the labor supply response of the working poor to the receipt of negative income tax payments; the emphasis on measuring the cost of national programs was really an afterthought. Spiegelman contended that the design proved relatively efficient for the original purpose; the variations in estimates across support levels and tax rates provided good measures of income and substitution effects. Second, in terms of the range of programs tested, it is important to note that training programs were added in some cases to counteract some of the adverse incentives. Third, the New Jersey experiment did serve as a feasibility study for later experiments, particularly Seattle-Denver. Fourth, the responses that the experiments were designed to measure were estimated with a fairly high degree of accuracy; despite the differences in sites, samples, and methodology, the labor supply response, particularly for males, fell in a fairly tight range across the experiments.

Harold Watts thought that Zellner and Rossi showed considerable naivete about how much time and money would be required to fulfill all the requirements of their textbook paradigm. The experiments tried to measure some basic behavioral responses and were quite successful in this regard. The results dramatically narrowed the range of estimates of the labor supply elasticities and this was a significant contribution to the debate. This conclusion seemed to reflect the consensus of the assembled group, albeit a somewhat biased sample since many had been involved in the design and execution of the experiments.

The Experiments in a Policy Context

Dennis Coyle and Aaron Wildavsky discussed the role of the income maintenance experiments in the gradual evolution of the negative income tax from an academic notion to a legislative proposal. Their paper focused specifically on the origins and ultimate defeat of President Nixon's Family Assistance Plan, and found that the preliminary results from the New Jersey income maintenance experiment had little influence on the final outcome. Instead, the authors attributed the failure of welfare reform in 1969-70 to the inability of representatives of different political cultures to achieve a compromise.

The negative income tax was endorsed in the 1960s by both liberals and conservatives in the wake of widespread disillusionment with the training and service programs of President Johnson's Great Society. When President Nixon came to office, he assembled a group of welfare experts to put together a domestic reform package that would eliminate poverty at a reasonable price. The result was the Family Assistance Plan, which would have provided to every family in the United States a minimum guaranteed annual income of $1600. The guaranteed income would have been reduced by 50 cents for each dollar earned by recipients until a break-even point of roughly $4000.

According to Coyle and Wildavsky, the specific design of the Family Assistance Plan was an attempt to appeal to three political cultures. The extension of benefits to millions of previously unprotected people without the stigma generally associated with welfare payments would please the "egalitarians," who support income redistribution. Limiting the plan to families would gain the backing of "hierarchs," who believe in the institution of the family and paternalistic social policies. Finally, letting the poor control their own expenditures would please the "individualists," who are committed to the autonomy of the individual.

In Coyle and Wildavsky's cultural notation, the public's attitude toward poverty at that time was a compound of hierarchy and individualism. Members of the public generally opposed a guaranteed income, preferring instead to guarantee and even require work. If poverty is the lack of money, the provision of money should end poverty. But if poverty is the lack of a job, and the discipline and self-respect that go with it, transferring money may only gloss over the poverty problem. It is better to give the poor what is good for them—food and work—which will enable them to be self-reliant and earn the individualist reward of the right to spend their earnings as they please.

The major view expressed in Congress about the Family Assistance Plan was that of the egalitarians, who reflected the attitude of the welfare establishment that the plan was essentially too little, too late. They repeatedly proposed alternatives that would broaden the definition of "family" to include all individuals and greatly raise the minimum income. Arguments that the Family Assistance Plan was a major step toward a universal guaranteed income failed to impress these liberal opponents. Eventually, the liberals united with conservatives, who reflected the public's belief that jobs, not money, held the answer to the poverty problem, and defeated the proposal.

The income maintenance experiments, originally designed to strengthen the case for a future negative income tax, became of immediate policy relevance when Nixon proposed reform along the lines of the New Jersey experiment. In response, officials of the Office of Economic Opportunity produced preliminary findings that indicated that work effort did not decline and may even have increased among those receiving payments. Although these results ran counter to economic theory, they were received enthusiastically by those supporting the bill. While later results showed that income guarantees reduced hours of work, the initial findings were still cited repeatedly by supporters of the negative income tax.

In any case, argued Coyle and Wildavsky, the experimental results were hardly equal to the task of overcoming fundamental cultural disagreements. In the end, the integrative solution embodied in the Family Assistance Plan — family support for hierarchs, extension of benefits for egalitarians, and reduced bureaucracy and greater autonomy for individualists — failed because adherents of these cultures refused to compromise. The egalitarians demanded a level of income guarantee unacceptable to individualists, while the hierarchs wanted to enforce values, especially a work requirement, that were unacceptable to either of the other cultures.

Lawrence Mead, the first formal discussant, had some sympathy with the authors' ideological approach, but attributed the failure of welfare reform in 1969-70 primarily to the fact that the politicians were out of step with public opinion. As repeated surveys indicate, the public wants to guarantee all needy persons subsistence, but wants to make the employable work for it. The reforming elites, however, were not willing to enforce social obligations in return for benefits.

Hugh Heclo argued that elaborate "cultural" theories were not necessary to explain the failure of welfare reform in 1969-70 and that the authors had failed to expose the important sociopolitical aspects of the income maintenance experiments. These experiments represented the triumph of an analytic subgovernment; no politician in the White House, no Congressman, no interest group as conventionally defined, and no lobby of ordinary citizens was pressing for multi-million-dollar social experiments. Their creation was the work of a more or less autonomous economics profession, which reflected both the growing prominence of economics and the relative collapse of its closest disciplinary competitor on poverty issues — social work/sociology. The dominance of the economists, however, meant that the experiments were very narrowly focused; Heclo characterized the exercise as "spending millions of dollars on four experiments to see if people worked less in response to income guarantees and next to nothing to find out what they did with any lessened time on the job."

The legacy of the experiments, according to Heclo, is twofold. In one sense, the experiments may have encouraged opponents of welfare reform to focus on the one issue of work incentives. On the other hand, the experiments broke ground for a whole succeeding generation of social experimentation. The new experiments employ more refined techniques and have closer connections to existing political and administrative structures. The history of social experimentation over the last 20 years must be admired as an attempt of a society to understand itself.

Policy Lessons and Implications for the Future

Members of a panel of experts, each from a different discipline, summarized their views about the policy lessons that resulted from the income maintenance experiments.

A Sociologist's Perspective

Lee Rainwater lamented that for all the money spent on the experiments, remarkably little was learned about social, as opposed to economic, behavior. He attributed this to three specific problems. The first was a lack of perspective in the initial conception of the experiments. The income maintenance experiments were designed only to test the implications of a negative income tax, which was a highly specific policy reflecting the particular circumstances of the time. Little thought was given to how this policy might fit into the range of available options, and almost no thought to how it might fit into the range of potential overall welfare regimes. Such a perspective might have been gained by looking at national policies in a comparative context; for example in Europe, economic security has always been linked to employment for working-age families.

Second, no effort was made in the experiments to penetrate the black box of causation. Few basic descriptive data were collected on what people thought was going on and why they reacted as they did. To do this would have challenged the basic tenets of modern social science, where the emphasis is placed on elegant manipulation of numbers rather than interpretation of narrative and qualitative information.

Third, because of the narrow focus of the study, the findings cannot tell us whether the negative income tax is good or bad policy. For example, an increase in the rate of marital separation and divorce (as initially claimed) need not be an undesirable development if people were dissolving destructive unions. Similarly, the reduction in work effort may not have adverse implications for a society with high levels of unemployment.

To Rainwater's list, commenter Charles Murray added three other reasons why the experiments failed to determine whether the negative income tax was good policy. First, no minimum baseline income standard exists that will enable everyone to have a decent standard of living. The conventional poverty index is meaningless, because it cannot discriminate between living a low-income life in the inner city and in a small town. A family at the poverty line might live decently in a civilized, functioning community, such as a small town in Missouri or Colorado, but be unable to survive on two or three times that amount in the South Bronx. Second, no one has considered what happens after a negative income tax is introduced nationwide and some people still have inadequate food and shelter; the merits of an income maintenance scheme that supplants the current system are very different from one that supplements it. Finally, the experiments were forced to focus on measurable outcomes and therefore provide no insights on noneconomic rewards, such as the psychic gains that people receive from earning their own income.

A Political Scientist's View

According to Richard Elmore, the experiments were designed to influence the political debate on income support in two ways. The first was methodological — to focus the debate on a few key empirical questions and estimate these effects more precisely than was possible with nonexperimental data — and the second was political — to legitimize the idea of a universal cash transfer program.

The main methodological lesson learned was that the very rigor of social experimentation limits the policy relevance of the results. The measured impact of the negative income tax on work effort would have to be qualified in a variety of ways to reflect the limited number of plans tested, the variability of results among different sites, misreporting of income and work, bias caused by attrition, variation in benefit packages available to control groups, and the difficulty of extrapolating from experimental results to a nationwide program. The alternative is to ignore the methodological uncertainties and average the results across experiments, but this approach undermines the methodological rationale for doing the experiments in the first place.

To the extent that the experiments have been successful as an instrument of political advocacy, their influence has been indirect. Although variants of the negative income tax found their way into the presidential or congressional arena five times, the published record shows that the experimental results entered the policy debate explicitly only twice. The first was the release of preliminary results from the New Jersey experiment in 1970 (discussed by Coyle and Wildavsky); the second occurred in 1978 when Senator Daniel Patrick Moynihan announced in a speech on the Senate floor that evidence of high rates of family dissolution among recipients in the Seattle-Denver experiment had caused him to question his earlier advocacy of a negative income tax. Neither of these instances captured the intent of policy researchers when they undertook the experiments. Moreover, the debate on the specific proposals focused very little on the estimates produced by the experiments. Rather, policymakers were more concerned with the incremental effects of changes in the design of the plans and with the winners and losers.

On the other hand, the analytic subgovernment that grew up around the experiments served as a place for stockpiling options, and when the problem-identifying and decisionmaking streams occasionally converged, these "option depots" supplied some of the raw material for the policy debate. Hence, research influences policy not by marshalling specific evidence in support of specific decisions, but rather by shaping policymakers' perceptions of the relevant policies and the feasible range of options.

Robert Reischauer argued that Elmore underrated the role of the experiments in legitimizing the negative income tax for policymakers; the findings were discussed frequently at meetings between congressional advocates of welfare reform and policy officials in the executive branch and they influenced the design of President Carter's welfare reform plan in numerous ways. Where the experiments failed was in convincing the American public that radical reform of the welfare system was necessary and desirable.

In Reischauer's opinion, failure was inevitable given that the negative income tax was designed to address the deficiencies that the policy elite saw in the current welfare system, not the shortcomings that most concerned the general public. The public believed that welfare costs were too high, that the caseload was expanding too rapidly, and that people who were fully capable of work were freeloading. In this setting, the experiments were bound to exacerbate the problem, because they focused on the measurement of labor supply responses to the proposed welfare reform. The results confirmed that indolence would be rewarded at the taxpayers' expense and thereby reinforced the public's negative perception of welfare reform.

An Economist's View

Robert Solow contended that social experimentation is bound to produce weak results—the coefficients are rarely statistically significant and the magnitudes of the responses are typically small. The nature of the results reflects both the inherent variability in each individual's behavior and the variation among individuals in their average response, which simply cannot be related to observed and observable characteristics. Nevertheless, social experiments may be useful in showing that policies selected on other criteria will not have dramatically destabilizing effects.

For example, economists embraced the negative income tax in the late 1960s because of the sense that the nation was finally in a position to eliminate poverty, the belief that the hodgepodge of categorical programs was inefficient, and the conviction that rules governing AFDC encourage family breakups. The one possible problem was that a decent guaranteed income combined with high tax rates required to keep costs under control would induce many recipients to withdraw from work. The experiments were designed to address this issue and they did produce an answer; guaranteed payments do have a labor supply effect, as economists predicted, but hardly large enough to jeopardize the nation's supply of work effort. Moreover, with continued high levels of national unemployment, the return of these individuals to the labor force probably would not have increased employment.

In Solow's view, the experience with the negative income tax provides a general model for social experimentation. Society may want to undertake certain policies for noneconomic reasons, but may be hindered by the fear that doing the right thing could be unexpectedly costly. A well-designed experiment can help determine the risks, and the prevalence of weak results should not be a deterrent.

Edward Gramlich thought that conference participants had been unduly critical of the experiments, pronouncing them a failure either because the research was inconclusive or because interest in the policy under investigation had waned. Disillusionment with the negative income tax, in his view, had nothing to do with the experiments, but rather reflected the need of taxpayers to be assured that responsibility for supporting the poor would be shared by recipients themselves, in the form of work requirements, child support enforcement, and other provisions that would have sounded punitive in the early 1970s. In Gramlich's opinion, the recognition of the need for responsibility sharing will eventually produce substantial welfare reform. The work-welfare experiments being carried out by the Manpower Demonstration Research Corporation, which have benefited technically and administratively from the negative income tax experiments, may have a positive impact on the nature of the reform, because they incorporate this element of responsibility sharing.

A Public Administrator's View

Barbara Blum addressed two questions. The first was one of process: What was the relationship between the way the income maintenance experiments were conducted and their reception by welfare officials? The second concerned substance: What lessons for administering today's welfare system were generated by the experiments?

Welfare administrators had little direct contact with the researchers who were conducting the experiments. One reason for the lack of communication was the difference in time perspectives of the two groups; the administrators were forced daily to confront a variety of new and pressing issues, while the researchers were engaged in an evaluation that would take several years to produce results. The nature of the particular experiments also created a gulf between the two groups. Researchers had little incentive to establish channels of communication with welfare administrators, who most likely would have been displaced if a negative income tax had been adopted. Hence, one problem associated with studying sweeping reform proposals is the difficulty of working closely with officials in the existing system to jointly identify and implement changes suggested by the research results.

Although the major findings of the experiments had no direct impact on the welfare system, some administrative procedures initiated by the researchers did find their way into existing programs. First, the researchers replaced the traditional procedure of infrequent face-to-face interviews to reevaluate eligibility with reports filled out and mailed in monthly by the recipients. Second, the researchers processed the reported data automatically. Third, they introduced retrospective budgeting so that benefits were based on the family's circumstances in the previous month, not on what it was anticipated they would need for the next one. Most states now use monthly reporting and retrospective budgeting, although some controversy exists about the effectiveness of these reforms with respect to both cost and the welfare of recipients.

Blum thought that two other interesting administrative issues were imbedded in the experiments. The first was the degree to which participants were actually aware of the rules of the game, since surveys indicated that only a fraction of beneficiaries understood how their benefits were calculated. Although analysts argue that people are better able to act in accordance with rules than to answer questions about them, the comprehension issue suggests that policymakers may defeat their purpose by making incentives so complex that rewards and penalties are obscured.

The second issue was whether it is desirable to have a more impersonal income maintenance system. For the many recipients who use welfare as a temporary source of aid, a simplified impersonal system would probably be highly desirable, and for this group it may be useful to look again at what was learned from the negative income tax experiments. But for chronic recipients, who consume a disproportionate share of the welfare dollars, it is probably necessary to provide a coordinated and sustained array of services in addition to benefit payments.

Wilbur Cohen did not consider the lack of contact between researchers and administrators a fatal flaw, since change is likely to be slow and incremental, as in the adoption of the administrative innovations. Future experimentation, however, should focus on modifying specific aspects of the current system, such as introducing work and training programs and determining the appropriate earnings disregard under AFDC.

Lessons for the Future

Richard Nathan summarized the lessons from the income maintenance experiments for both social policy and future research. In his opinion, the main effect on social policy was to educate government officials, the media, and interested citizens on the issues associated with the introduction of a negative income tax. The educational process was expensive and also cast doubt on the idea as a solution to the nation's poverty problem. Giving money to people without requiring work, however, was never a comfortable approach for most politicians, and for this reason Nathan concluded that the negative income tax was an ill-advised subject for social experimentation. Experiments should be restricted to situations where the politicians are "(1) genuinely interested in dealing with an issue; (2) uncertain about how to do so; and (3) willing to consider the approach that is the subject of experimentation." The negative income tax did not satisfy these conditions.

In terms of policy research, the experiments demonstrated that it was possible to conduct large-scale, rigorous, honest demonstration projects with random assignment of participants to treatment and control groups. On the other hand, since social experiments are expensive and take a long time to complete, researchers should attempt to learn more from such endeavors than they did in the negative income tax case. Nathan also argued that experiments of more selective service-type initiatives are to be preferred over demonstrations of universal transfer schemes. Not only are such policies more realistic politically, but the results of such experiments are more easily applied to the nation as a whole, whereas introducing a massive income transfer scheme might change national behavior in unforeseeable ways.

In short, Nathan concluded that while the negative income tax experiments were unwise, the idea of social experimentation with random assignment, which they introduced, is good. "The negative income tax experiments, as the first such effort of this type, led the way in developing both the capacity and the sensitivity necessary to the more effective use of social experimentation as an input to the government process."

Conclusions

In terms of an overall assessment of the income maintenance experiments, the conference participants fell into two groups. One argued that the effort absorbed an inordinate amount of the available research funds and diverted professionals from other, more worthy endeavors. The other contended that the experiments were a useful device that not only improved the existing estimates of labor supply responses but also increased our capacity to carry out social science research.

The debate over whether the experiments were worthwhile in view of the opportunities forgone will never be resolved, but almost all experts agree that two important results emerged. First, the experiments refined the estimates of individuals' responses to net wage rates, measured by using variations in taxes, and to unearned income, demonstrated by using variations in guaranteed income. The results of the income maintenance experiments are valuable not only for evaluating the effects of welfare reforms, but also for estimating the effects of changes in other programs, such as expanding the earned income tax credit in the personal income tax. Moreover, even though attention has now turned to programs that will require work for welfare benefits, the estimates are useful to show the parameters that the administrators are pushing against.

The second lesson from the experiments, namely the merits of random assignment, is even more important if Congress endorses the Administration's proposal to embark on a series of state experiments in welfare reform. If these experiments are to help in improving the welfare system, they must assign participants randomly to control and treatment groups. Only this approach avoids self-selection bias, a phenomenon for which no statistical method can compensate. Nowhere are the difficulties of evaluating programs without random assignment more apparent than in Massachusetts. Encouraging results have been claimed for the state's Employment and Training (ET) Choices program, but the lack of a control group makes it impossible to separate the effects of the training program from the impact of an economy operating with very low levels of unemployment.

Recent social experimentation has demonstrated its ability to produce timely results at a reasonable cost. It would be criminal for the states to spend the next decade experimenting with a host of alternative approaches to welfare reform without providing the bases for evaluating them.

The Work Response to a Guaranteed Income: A Survey of Experimental Evidence

*Gary Burtless**

Presidents and policy analysts are periodically seized with a passion to reform the nation's welfare system. This passion occasionally results in a serious proposal for thorough reform, such as President Nixon's Family Assistance Plan, President Carter's Program for Better Jobs and Income, or President Reagan's New Federalism. The only reform proposal that has received experimental scrutiny, however, is a suggestion advanced by academic economists—the negative income tax or guaranteed annual income plan. While popular among economists, the negative income tax proposal has never attracted much enthusiasm—or even attention—among politicians and voters. Fortunately, the findings from the negative income tax experiments are relevant to a wide variety of reform proposals, including the plans suggested by recent Presidents. Experimental results were used, in fact, to predict the behavioral consequences of both the Nixon and the Carter reform proposals. This essay summarizes the labor supply findings from the four negative income tax experiments and considers their implications for reforming the American welfare system.

It is useful at the start to distinguish among three different kinds of labor supply estimates that have been produced by the experiments. The first was obtained by measuring the simple difference between the work effort of people who were assigned to experimental negative income tax plans and that of people who were assigned to the control group. Those in the control group were not eligible to receive payments

*Senior Fellow, The Brookings Institution. David Betson, Robert Moffitt, and Philip Robins provided useful suggestions in the preparation of this paper. In addition to the two discussants, David Greenberg, Robert Haveman, Robert Reischauer, and Alice Rivlin gave helpful comments on an earlier version. The opinions are the author's own and should not be ascribed to the staff or trustees of The Brookings Institution.

and so were presumably unaffected by the experiment. The labor supply difference between these two groups is ordinarily expressed as a reduction in average hours of work per week or as a percentage change in comparison to the average hours worked by members of the control group.

A second type of estimate is produced by using structural models of work effort response. Structural models yield a decomposition of the overall work reduction into a change that is due to the net wage or tax change, on the one hand, and one that is due to the increase in family income, on the other. These two separate effects are referred to as the substitution and income effects, respectively. Economists usually prefer this type of measure of response for two reasons. It permits the results of the experiments to be directly compared with labor supply findings from nonexperimental studies. And it allows analysts to generalize the findings from the experiments to a much wider population than the one enrolled in the experiment and to estimate the effects of a broader range of plans than the ones actually tested in the experiment. The latter advantage is particularly important from the point of view of evaluating realistic reform proposals, since no plan suggested by congressmen or Presidents has borne much resemblance to the ones tested in the experiments.

The third kind of estimate of response is generated using microeconomic simulation. This type of estimate is simply a generalization of the results from an experiment to the national population. Using estimates of the income and substitution effects obtained in an experiment and a microeconomic census file representing all U.S. households, economists have predicted the response of low-income workers to alternative income maintenance tax plans and summed these responses to produce an estimate of the effect on national labor supply.

From a scientific viewpoint the most reliable estimate of work reduction is the simple difference in labor supply between members of the treatment and control groups. This is the measure of response that the experiments were specifically designed to produce and it is the one that has been most widely reported in the popular press. It is inherently more difficult to decompose the overall response into income and substitution effects, although in this respect the experiments possess substantial advantages over nonexperimental sources of data. The experimentally based simulations of national response are more problematical. National simulations are based on a specific (and perhaps erroneous) decomposition of the experimental response into income and substitution effects and on detailed assumptions about the responses of subpopulations that were unrepresented or poorly represented in the experiments.

Unfortunately, from the perspective of their policy usefulness, the

Table 1
Description of the Negative Income Tax Experiments

	Characteristics of the sample				Characteristics of the plans			
Experiment	Sample Size	Family Composition	Race	Income Truncation[a]	Duration	Range of Guarantee[a]	Range of Tax rates (percent)	Range of Breakeven[a]
New Jersey (1968-1972)	1,357	Husband-wife (100%)	White (32%) Black (37%) Hispanic (31%)	150	3 years	50 to 125	30 to 70	100 to 250
Rural (1969-1973)	809	Husband-wife (85%) Single female parent (15%)	White (65%) Black (35%)	150	3 years	50 to 100	30 to 70	100 to 250
Gary (1971-1974)	1,780	Husband-wife (41%) Single female parent (59%)	Black (100%)	None[b]	3 years	77 and 101	40 and 60	128 to 253
Seattle-Denver (1971-1982)	4,800	Husband-wife (61%) Single female parent (39%)	White (39%) Black (43%) Hispanic (18%)	325	3 years (71%) 5 years (25%) 20 years (4%)	92 to 135	50 and 70 $70 - .0025Y$[c] $80 - .0025Y$[c]	140 to 300

[a] Measured as a percent of the poverty line. Breakeven is the income level at which the negative income tax payment is reduced to zero. Partial reimbursement of income and payroll taxes was phased out at higher income levels.

[b] The Gary sample was initially restricted to families with incomes below 240% of the poverty level, but a small sample with incomes above this limit was subsequently enrolled to minimize truncation bias.

[c] Declining marginal tax rate plans. Y is family income, implying that the marginal tax rate declined by 2.5 percentage points with every $1,000 increase in income.

Sources: Committee on Finance, U.S. Senate (1978), p. 316, Kehrer (1977), and Robins (1985).

three kinds of estimates of response would rank in the reverse order. The average difference in labor supply between treatment and control groups within a particular experiment may be suggestive, but it is not especially helpful for predicting the effect of a realistic welfare reform plan on a representative population. The most useful and meaningful estimates of response are ones that reflect the response of a nationally representative sample to a plausible program of reform. Unfortunately, such estimates are inherently the least reliable.

The remainder of this paper considers, in turn, the three kinds of estimates just described and their major limitations. The paper concludes with a discussion of the implications of the estimates for welfare policymaking.

Simple Estimates of Response

The negative income tax experiments produced a large number of estimates of average response to the tested plans. These estimates naturally vary across the four experiments, since the experiments tested different plans on different populations. Table 1 describes some of the main features of the samples and negative income tax plans tested in the experiments.

The samples varied tremendously in the different experiments. The first experiment, in New Jersey and Pennsylvania, enrolled low-income black, white, and Hispanic residents of declining urban areas. All of the enrolled families originally contained both husband and wife. The Rural experiment, which was conducted in Iowa and North Carolina, contained low-income rural families. Although most of the families contained both a husband and a wife, a small number of single-parent families were also enrolled. The two later experiments, in Gary, Indiana, and in Seattle and Denver, enrolled higher income samples drawn from low-income census tracts in large midwestern and western central cities. The samples in these experiments were purposefully drawn to represent single-parent as well as two-parent families. The Gary experiment was restricted to black families, while the Seattle-Denver experiment included large samples of white and Chicano, as well as black, families. Clearly, the differences in the samples are important enough so that significant differences might be expected in the average response even if each of the experiments had tested an identical set of plans.

The tested negative income tax plans were not identical, however. On average, the New Jersey, Rural, and Gary experiments tested less generous plans than the ones tried in Seattle and Denver. That is, the Seattle-Denver plans offered more generous payments to families without other income and provided payments to families at higher in-

Table 2
Changes in Hours and Earnings in Four Negative Income Tax Experiments
(Percentage changes in parentheses)

	Husbands		Wives		Single female heads of families	
Experiment	Hours per year	Annual earnings[a]	Hours per year	Annual earnings[a]	Hours per year	Annual earnings[a]
New Jersey						
White	−99 (−5.6)	+10 (+0.1)	−73 (−30.6)	−420 (−33.2)	—	—
Black	+36 (+2.3)	+1,180 (+9.3)	−5 (−2.2)	+110 (+7.8)	—	—
Hispanic	−10 (−0.7)	+800 (+6.4)	−99 (−55.4)	−560 (−54.7)	—	—
All[b]	−21 (−1.2)	+690 (+5.3)	−56 (−24.6)	−270 (−21.4)	—	—
Rural[c]						
White	+40 (+1.8)	−590 (−4.8)	−88 (−21.1)	−170 (−12.1)	—	—
Black	−152 (−8.0)	−630 (−6.8)	−268 (−31.3)	−1,360 (−41.6)	—	—
All	−56 (−2.8)	−610 (−5.7)	−178 (−27.9)	−770 (−32.8)	—	—
Gary						
Black	−114 (−6.5)	−830 (−5.0)	+14 (+5.0)	+160 (+10.5)	−112 (−30.0)	−280 (−13.9)
Seattle-Denver						
White	−144 (−7.6)	−1,310 (−7.5)	−107 (−17.1)	−590 (−16.5)	−85 (−8.6)	−900 (−13.9)
Black	−169 (−9.5)	−930 (−5.9)	−153 (−16.0)	−860 (−15.6)	−180 (−16.6)	−980 (−14.0)
Hispanic	−231 (−11.5)	−510 (−3.0)	−147 (−28.7)	−800 (−32.5)	−202 (−20.4)	−1,380 (−22.3)
All	−164 (−8.8)	−1,070 (−6.4)	−128 (−17.9)	−710 (−17.6)	−144 (−14.0)	−1,000 (−14.9)
3-year Sample	−133 (−7.1)	−810 (−4.8)	−101 (−14.2)	−580 (−14.4)	−134 (−13.0)	−940 (−14.1)
Weighted average[d]	−119 (−7.0)	−650 (−4.0)	−93 (−17.0)	−480 (−16.0)	−133 (−17.0)	−760 (−15.0)

[a] Annual earnings changes are measured in 1985 dollars. Earnings estimates reported in original reports were converted using the personal consumption expenditure deflator.

[b] Results for overall New Jersey response are obtained by weighting responses of separate racial groups. Racial weights are reported in Table 1.

[c] Results for Rural response are obtained by weighting of separately reported responses for white wage earners in Iowa (25% of sample), white wage earners in North Carolina (25%), and black wage earners in North Carolina (50%).

[d] Separate responses are weighted using reported estimation samples in four experiments. For husbands and wives, New Jersey = 0.20; Rural = 0.07; Gary = 0.17; Seattle-Denver = 0.56. For female heads, Gary = 0.34, and Seattle-Denver = 0.66.

Sources: New Jersey: Rees (1974), pp. 174-75; Rural: U.S. Department of Health, Education and Welfare (1976), pp. 23 and 29; Gary: Moffitt (1979), p. 482, and Greenberg, Moffitt, and Friedman (1981), p. 586; Seattle-Denver: SRI International (1983), pp. 120-22 (second experimental year results) and Robins and West (1980b), pp. 16, 19, 22, and 59-67.

come levels. Thus, other things equal, we would expect the Seattle-Denver plans to induce a larger response. Other things were not equal, however. I have already mentioned differences in the income distributions of the four samples. In addition, the nonexperimental welfare benefits available to members of the control group differed across the experiments. Local labor market conditions also differed. There is thus no reason to expect that the average response to the income maintenance plans would be identical across experiments.

Table 2 shows the average work effort and earnings reductions within various subsamples of the four experiments. The estimates are taken from the final reports of each of the experiments. Analysts essentially estimated a statistical model of the following type:

$$Y = \alpha + \beta T + \gamma Z + \delta X + \epsilon, \tag{1}$$

where Y is the dependent variable of interest (either hours of work or earnings), T is a treatment dummy variable that takes the value one for people assigned to any of the negative income tax plans and zero for members of the control group, Z is a vector of variables originally used to stratify the sample in the experimental design (for example, pre-experimental income level), and X is a set of personal characteristics believed to affect the dependent variable (age, educational attainment, place of residence, and so forth). The treatment effect is β, and it captures the average effect of treatment on an average member of the sample assigned to negative income tax plans.[1]

Most but not all of the entries in table 2 are negative, implying that the negative income tax plans caused reductions in work effort and earnings for most subsamples enrolled in the experiments. All of the entries for the Seattle-Denver experiment are negative and, especially for men, are often larger than corresponding entries from the other three experiments. Virtually all of the Seattle-Denver estimates are significantly different from zero at the 95 percent confidence level, whereas estimates from the other experiments are frequently insignificant. There are two explanations for this pattern. As mentioned above, the average generosity of the Seattle-Denver plans was greater than that of the plans tested in the other experiments, causing a larger response, and the sample enrolled in Seattle-Denver was much larger, yielding a smaller standard error around the point estimate of response.

The bottom row in table 2 shows average hours and earnings reductions in all of the experiments for husbands and wives in two-parent families and female heads of single-parent families. Husbands reduced their reported work effort by approximately 7 percent, while wives and female heads reduced reported hours by 17 percent. The greater responsiveness of women than of men is consistent with the relative labor supply elasticities reported in nonexperimental studies. Analysts of the Gary and Seattle-Denver experiments concluded that most of the hours reduction was caused by shorter durations of employment and by longer durations of unemployment and labor force withdrawal among people enrolled in the negative income tax plans.[2] There was only a comparatively small effect on the weekly hours of those remaining at work.

On balance, the proportional reductions in earnings were quite close to the reductions in hours. Although the earnings reductions might appear to be relatively modest, they are sizable when compared with the negative income tax payment received by a typical family. In the Seattle-Denver experiment, for example, eligible two-parent families received transfer payments that were $2,700 larger than the nonexperimental payments sent to members of the control group.[3] The combined earnings reduction of husbands and wives in the Seattle-Denver treatment group was almost $1,800, or approximately two-thirds of the net experimental payment. The average tax rate of the Seattle-Denver plans was about 50 percent, implying that the $1,800 earnings reduction caused payments to be $900 above what they would have been in the absence of a work effort response. Thus, one-third of the net transfer cost of the Seattle-Denver plan was due to the reductions in reported earnings among participants. Another way to interpret the same set of figures is to say that the experiment spent nearly $2,700 on transfers and succeeded in raising the incomes of two-parent families by only $900.[4] Even if the earnings reductions are taken to be modest, it is reasonable to ask whether most taxpayers would be willing to spend $3 in order to raise the incomes of poor, two-parent families by only $1.

Several analysts have found evidence that at least part of the employment and earnings reduction reported in the experiments was spurious. Recipients of negative income tax payments had a clear incentive to underreport their employment and earnings, because to do so permitted them to receive a larger payment than the one to which they were legally entitled. Wage earners enrolled in the control group did not face this kind of misreporting incentive.

It is possible to analyze this issue with sufficiently accurate employment and earnings data which are not subject to reporting bias. The employment and earnings records of the unemployment insurance system provide one source of such data. The effects of underreporting were systematically examined using these data in two of the experiments, Gary and Seattle-Denver. In both experiments underreporting was found to bias the estimates of employment and earnings response. The bias in the Gary experiment was large enough so that the entire earnings response and much of the apparent employment effect of the experiment disappeared.[5] In the case of the Seattle-Denver experiment the bias was somewhat smaller and less reliably estimated. Husbands and women heading single-parent families misreported their employment and earnings infrequently enough so that the response estimates reported in table 2 are probably only slightly overstated. On the other hand, the responses of wives and other secondary earners are greatly overstated.[6] The earnings reduction of wives, for example, virtually disappears when the response estimate is based on presumably accurate data from the unemployment insurance system. Of course, even if misreporting bias causes an exaggeration of the efficiency loss from a negative income tax, there is no reason for complacency about the earnings reductions reported in table 2. An earnings reduction caused by underreporting is just as costly to taxpayers as a reduction caused by a genuine reduction in work effort.

Offsetting the bias from misreporting is the effect of the limited duration of the experiments. There are at least two reasons to believe that a limited duration income maintenance program will elicit a smaller response than a permanent program that offers the same income guarantee and tax rate. The first is that workers may need time to respond to the incentives embedded in an income maintenance plan. If they are given only three years to respond, as they were in the experiments, their eventual response might not be fully observed. A second reason to expect a small response is that the income effect produced by a limited-duration program is by definition less than the income effect produced by an otherwise equivalent program which is expected to be permanent. A $1,000-per-year payment should cause a larger effect if it is to last indefinitely than if it is to continue only three years, unless the worker applies an extremely high discount rate to future income. On the other hand, because the experiments were temporary they essentially offered a sale on leisure, which participants were forced to take advantage of within a concentrated period. This encouraged greater responsiveness than would have been observed in a permanent program.

The Seattle-Denver experiment is the only one that permits us to examine the effects of limited duration in a reasonably satisfactory way. About 30 percent of the eligible sample in that experiment was enrolled for five years, while the remainder was enrolled for three years.[7] The pattern of response of families in both the five-year and three-year groups suggests that workers were somewhat slow in reacting to the negative income tax disincentives. Robins and West (1980b, p. 36) estimate that 90 percent of the full response would only be observed after 2.4 years in the case of husbands, after 3.6 years in the case of wives, and after 4.5 years in the case of single women with dependent children.

The same authors also find that the average response of husbands and wives (though not of single women) in the five-year treatment group was substantially greater than that in the three-year group, even when the responses of the two groups are measured at the same point in time (for example, two years after enrollment). The maximum response of husbands in the five-year group occurred in the third and fourth years of the experiment, when the hours reduction was 13 percent and the earnings reduction approximately 12 percent (Robins and West, 1980b, p. 23). These reductions are about twice the magnitude of responses in the three-year group during the second year. The maximum response for wives in the five-year group occurred in the fourth and fifth years, when the hours reduction was 27 percent and the earnings reduction 26 percent (Robins and West, 1980b, p. 25). For single women heading families, the maximum response occurred in the fifth year, when the hours reduction was 32 percent and the earnings reduction about 35 percent (Robins and West, 1980b, p. 27).

Clearly, the sluggishness of the labor supply response and the attenuated response to a shorter duration plan cause the long-term impact of a permanent negative income tax to be substantially understated by the mid-experimental responses of families assigned to three-year plans. It should be stressed, however, that this conclusion is valid only for the relatively generous, low-tax plans tested in Seattle-Denver. Plans with high tax rates might have elicited a different pattern of response. Burtless and Greenberg (1982) found evidence that participants in the experiment, particularly women, reacted more strongly to the tax rates than they would have in a program of permanent duration. Taking advantage of the sale on leisure, participants in the three-year plans were significantly more responsive to the tax than were participants in the five-year plans. If a high-tax, low-guarantee plan had been tested, it is conceivable that the overall response of the three-year treatment group would have been larger, not smaller, than that of the five-year group.

The implications of table 2 may be summarized briefly. The four negative income tax experiments caused moderate to large proportional reductions in work effort. As expected, the proportional response was greater among women than men. The absolute reductions were largest in the Seattle-Denver experiment, which offered the most generous plans, and were smaller and less precisely estimated in the experiments testing plans with a lower income guarantee and breakeven point. The work effort reductions were overstated due to misreporting bias but understated because of the limited duration of the experiments, particularly in the case of high-guarantee, low-tax-rate plans. On balance, the experiments probably underestimated the permanent response to a negative income tax program with a generous guarantee (equal, for example, to the poverty line) and a relatively low tax rate (equal to or below 50 percent). It is less certain that the effect of low-guarantee, high-tax negative income tax plans would be understated in a short-duration experiment.

Even if we had perfect confidence in our estimates of average response, it is not clear how they could be used to predict the consequences of reform plans in which policymakers are actually interested. The Seattle-Denver experiment produced the most precise results and the ones that have been subject to the most thorough sensitivity analysis. But those results were obtained in an experiment in which the average negative income tax plan provided a guaranteed income of about 115 percent of the poverty line and taxed earned income at a marginal rate of approximately 50 percent. No feasible welfare reform plan could offer universal benefits this generous. The maximum combined benefit from aid to families with dependent children (AFDC) and food stamps is now only 73 percent of the poverty line in the median state and is equal to the poverty line in only the most generous state. The gross income limit for receiving benefits from AFDC is less than twice the poverty line in all but two states and below 1.5 times the poverty line in 37 states.[8] By contrast, over half of the families in the Seattle-Denver experiment were enrolled in plans with an income breakeven above twice the poverty level, and 86 percent were enrolled in plans with a breakeven above 1.5 times the poverty line.[9] Thus, not only were transfers quite generous in the Seattle-Denver experiment, they were available to families well up in the income distribution. For these reasons, the average recorded response to the Seattle-Denver plans does not provide a useful approximation of the expected response to plausible programs of welfare reform.

Structural Models of Response

Because estimates of the average negative income tax response are difficult to use, economists analyzing the experiments have sought to obtain structural estimates of response. We can distinguish between two broad classes of structural models. The first emphasizes the response within an experiment to the separate negative income tax plan characteristics—the income guarantee and the tax rate. The second emphasizes the individual-level response to unearned income and net wage levels more generally, or to changes in these two variables induced by the experiment.

At first blush, the first type of estimate might appear to be the easiest to produce and then apply in predicting the effects of income maintenance alternatives. To estimate the separate responses to income guarantees and tax rates, the analyst simply estimates equation (1) but replaces the variable T with a set of variables that reflect the level of the experimental guarantee, the marginal tax rate, and possibly some interaction between these two program features. For example, if the experiment tested a low, moderate, and high guarantee and a low, moderate, and high tax, with complete interaction of the guarantee and tax, a simple way to represent the treatment is:

$$(2) \qquad H = \alpha + \beta_1 T + \beta_2 (\Delta G) + \beta_3 (\Delta t) + \gamma Z + \delta X + \epsilon,$$

where T is again a variable for assignment to experimental treatment, ΔG is the dollar difference between the individual's assigned guarantee and the lowest guarantee tested, and Δt is the percentage-point difference between the assigned tax rate and the lowest tax rate tested. (ΔG and Δt both take the value zero for control observations and individuals assigned to the low-guarantee, low-tax plan.) With this specification, β_1 is interpreted as the average effect of the low-guarantee, low-tax program, β_2 is the average effect of a one-dollar rise in guarantee, and β_3 is the effect of a one-point rise in the tax rate. More complicated interaction effects of the guarantee and tax can also be specified. The expected effects of an alternative negative income tax plan can be predicted using the estimates of β_1, β_2, and β_3 and suitably defining T, ΔG, and Δt to represent accurately the alternative plan.

This approach to estimation, although straightforward, is less useful than it first appears. The samples enrolled in the experiments were not nationally representative, so the estimates of β_1, β_2, and β_3 will not necessarily be valid if applied to a wider population. For example, the average effect of a low-guarantee, low-tax plan, β_1, depends on the generosity of the welfare system against which it is compared. The plan might be expected to cause little work reduction in a state like Washington, where AFDC benefits are high, but significant reductions in Indiana, where the maximum payment for such aid is extremely low. Since the experiments did not enroll samples that faced a representative set of state welfare programs, it is not clear how estimates of β_1, β_2, and β_3 can be used to predict work effort responses in states where no experiment was conducted.

A more subtle problem arises because of the sampling design used to assign families to negative income tax treatments. The experiments did not use simple random assignment. The potential sample in each experiment was divided into subsamples defined by a set of stratifying variables. One important stratifying variable was preexperimental income level. The Seattle-Denver experiment, for example, divided the sample into seven preexperimental income classes. Families within each income class were randomly assigned to one of the tested plans or to control status. The proportion assigned to a specific plan was not identical in each income class, however. In order to increase the number of families that could be enrolled given a fixed budget constraint, the experiments enrolled a higher proportion of low-income families into the least generous plans and a higher proportion of high-income families into the most generous plans. The income distribution of families assigned to the most generous plans was consequently not the same as that of families assigned to the least generous plans. This implies that the differences in average work effort response to two different plans in the same experiment may be due to differences in the composition of the samples assigned to the plans as well as to genuine differences in response induced by the plans.[10] This problem could be avoided in estimation by fully interacting the negative income tax plan parameters with the stratifying variables, but such a procedure is extremely cumbersome and yields statistically imprecise results.[11] No published study from the experiments relies on this approach. When analysts have estimated structural models of response to the income guarantee and tax rate, they have not estimated all of the interaction terms that would permit us to disentangle the effect of the sampling plan from that of the treatments themselves.

In a second approach to estimation, economists have specified labor supply models quite similar to those estimated with nonexperimental data. A model of this type was estimated by Keeley et al. (1978b) using data from the Seattle-Denver experiment:

$$(3) \qquad H = \alpha + \beta_1 \, [\Delta w|_{\bar{H}_P}] + \beta_2 \, [\Delta Y|_{\bar{H}_P}] + \gamma Z + \delta X + \epsilon,$$

where Δw is the change in the after-tax wage rate caused by an individual's assigned negative income tax plan and ΔY is the change in after-tax income. Both Δw and ΔY are computed at the individual's preexperimental level of work effort, H_P. Obviously, Δw and ΔY will vary widely even among individuals facing the same negative income tax plan. Under the usual assumptions we would expect β_1 to be positive and β_2 to be negative. That is, increases in the net wage, holding income constant, should cause individual labor supply to rise, while increases in income, holding the net wage constant, should reduce labor supply. The Keeley et al. specification is similar to one estimated with nonexperimental data and proposed by Ashenfelter and Heckman (1974). Like Keeley et al., Burtless and Hausman (1978) estimated a model that could be applied as easily to nonexperimental as to experimental data:

$$(4) \qquad \log (H) = \alpha + \beta_1 \, w + \beta_2 N + \delta X + \epsilon,$$

where w is the after-tax wage rate and N is the virtual income intercept

Table 3
Estimates of Substitution and Income Elasticities from Experimental and Nonexperimental Studies

Subjects		Uncompensated Substitution Elasticity[a]	Compensated Substitution Elasticity[b]	Total Income Elasticity
Men				
Negative Income Tax	[N = 21][c]	.0043	.0795	−.0757
		(.098)	(.068)	(.093)
Weighted Negative Income Tax		−.0223	.0902	−.1139
Nonexperimental	[N = 26][c]	−.1045	.2842	−.3873
		(.178)	(.415)	(.339)
Women				
Negative Income Tax Wives	[N = 20][c]	−.0420	.1105	−.1515
		(.368)	(.237)	(.214)
Weighted Negative Income Tax		.0659	.1783	−.1115
Negative Income Tax Wives[d]	[N = 14][c]	.0957	.1907	−.0957
		(.225)	(.154)	(.146)
Weighted Negative Income Tax		.1730	.2425	−.0696
Negative Income Tax Female Heads	[N = 11][c]	−.0373	.1346	−.1709
		(.123)	(.070)	(.085)
Weighted Negative Income Tax		−.0426	.1355	−.1774
Nonexperimental	[N = 48][c]	1.9919	2.0248	−.0331
		(3.162)	(3.154)	(.423)
Nonexperimental[e]	[N = 38][c]	1.3553	1.3661	−.0113
		(1.319)	(1.229)	(.463)

Numbers in parentheses are standard deviations.

[a] $\left(\frac{\Delta H}{\Delta W(1\text{-}t)}\right) \div \left(\frac{H}{W(1\text{-}t)}\right)$

[b] $\left(\frac{\Delta H}{\Delta W(1\text{-}t)}\bigg|_{\bar{U}}\right) \div \left(\frac{H}{W(1\text{-}t)}\right)$

[c] Number of separate estimates of response used to compute the reported elasticity.

[d] Excludes estimates from the New Jersey experiment.

[e] Excludes five estimates showing the highest compensated substitution elasticity and five estimates showing the lowest compensated substitution elasticity.

Sources: See text.

measured at an individual's desired hours of work.[12] For low-wage workers, it is reasonable to expect β_1 to be positive and β_2 to be negative.

A cynic might ask why it is necessary to invest $100 million collecting experimental data when analysts then estimate models that could as easily be estimated using nonexperimental data. While the question is a legitimate one, it has a straightforward answer. The experimental variation in tax rates and income guarantees produces a large amount of essentially random variation in Δw and ΔY and in w and N, the critical variables in equations (3) and (4). The variation is not totally random, of course, because these variables are correlated with a worker's gross wage rate and may be correlated with preexperimental work effort and other confounding variables through the effects of the negative income tax plan assignment procedure, discussed above.[13] But in spite of this correlation, random assignment of workers to widely differing negative income tax plans assures us that a greater fraction of the variation in Δw, ΔY, w, and N will be independent of observed and unobserved variables that affect H.[14] From a statistical standpoint, this should increase our confidence in the resulting coefficient estimates.

A large number of structural models have been estimated using data from the negative income tax experiments, particularly the Seattle-Denver experiment. Moffit and Kehrer (1981, pp. 138-42) and Robins (1985, p. 578) have reported individual and average estimates of income and substitution effects obtained in each of the negative income tax experiments. Table 3 presents a summary of income and substitution elasticities, averaged across the four experiments. These estimates of response in the New Jersey, Rural, and Gary experiments are based upon corrected estimates of elasticities reported by Moffitt and Kehrer (1981). Estimates for the Seattle-Denver experiment are based on the simple average of elasticities reported in seven separate studies using data from that experiment.[15] Table 3 shows the average experimental estimates of the (uncompensated) net wage elasticity, the compensated substitution elasticity, and the total income elasticity for husbands, wives, and female household heads. The substitution elasticities are useful in indicating the slope of the labor supply function and the rough magnitude of efficiency losses arising from imposition of higher tax rates on low-wage workers. The total income elasticity shows the percentage by which work effort falls with a one percentage point rise in income that is not accompanied by a change in the net wage rate.

Average elasticity estimates for the experiments were computed in two different ways. First, the simple arithmetic average of all of the estimates from separate studies of labor supply response was calculated. The top row in the table, for example, shows the simple average of 21 separate estimates of the labor supply elasticity for husbands enrolled in the experiments.[16] In parentheses below these elasticity averages the table shows the standard deviation of the different point estimates of response around the average estimate. A more defensible way to compute the mean response is to account for the relative size of the samples used to obtain different elasticity estimates. It seems reasonable, for example, to attach a lower weight to the estimated elasticity within a sample of 200 New Jersey Hispanics than we attach to the response of 2,200 white, black, and Hispanic husbands in Seattle and Denver. A weighted estimate of the average elasticity was derived using a two-step procedure. First the average elasticity within each of the four negative income tax experiments was computed, and then the weighted average elasticity was calculated by suitably weighting the measured responses in the four experiments. (Weights are reported in a footnote to table 2. Where necessary the responses of separate racial groups within each experiment were weighted. See the weights reported in table 1.)

It might be argued that the estimates of response from the individual studies should be weighted by the quality of the research methodology rather than the size of the estimation sample. This is the implicit strategy of Borjas and Heckman (1978) in an early survey of the nonexperimental labor supply literature. Such a survey would yield more interesting and precise results than those reported here. However, it would also require thorough justification of the weights attached to the various studies. I will leave that exercise to others.

It is useful to compare the estimates obtained in the experiments with labor supply estimates reported in the nonexperimental literature. Killingsworth (1983) has provided an informative survey of elasticity estimates obtained in nonexperimental studies. Table 3 contains my computations of the average and standard deviation of elasticity estimates reported in 26 nonexperimental studies of U.S. prime-aged men and 48 studies of U.S. women.[17] Because the range of estimates for women was so large, average female elasticities were computed excluding the five studies with the highest and the five studies with the lowest estimates of compensated substitution elasticity.

The labor supply functions estimated with experimental data appear to be comparatively inelastic. For example, the uncompensated labor supply function of low-wage men is essentially vertical. A change in the net wage, holding nonwage income constant, has virtually no effect on annual male work effort. Even if we consider estimates of the uncompensated elasticity one standard deviation from the mean estimate, the elasticity appears to be quite moderate.

The uncompensated substitution elasticity of wives is less reliably estimated. Although the average estimated elasticity is only -0.04, this average is sensitive to the method of weighting. When the several elasticity estimates are weighted according to the size of the estimation samples, the mean elasticity rises to $+0.07$. Much of the uncertainty arises because of the lack of robustness of estimates of wives' supply elasticities in the New Jersey experiment. When the New Jersey estimates are excluded, the mean uncompensated elasticity rises to 0.10–0.17, depending on how the remaining estimates are weighted. Note that the standard deviation around the unweighted average falls by more than one-third when New Jersey estimates are excluded. The relatively large dispersion in estimates of the labor supply of wives was caused by the income truncation imposed on samples enrolled in the experiments. Since the samples were restricted to very low-income families, they contained an abnormally small percentage of working wives. In the experiments with the lowest income limits (New Jersey and Iowa-North Carolina), the elasticity estimates were sensitive to the work effort changes of only a handful of women. The elasticity estimates for men and for women heading single-parent families seem to fall in a much narrower range than do the estimates for wives.

In comparison to the estimates from the nonexperimental literature, the elasticity estimates from the experiments tend to be much smaller in absolute value. This tendency is most pronounced with respect to the compensated substitution and the income elasticities estimates for men and, even more strikingly, for the uncompensated and compensated substitution elasticities for women. Whereas most nonexperimental estimates show a strongly positive uncompensated supply function for women, the experiments found only weakly positive or even backward-bending supply functions. The mean experimental estimates of the income elasticity for men and women are in the range −0.07 to −0.18. These estimates are below the average nonexperimental estimates in the case of men but above the average nonexperimental estimates for women.

On balance, the experimental estimates imply a smaller responsiveness to negative income tax disincentives than do most nonexperimental estimates. This conclusion was also reached by Moffitt and Kehrer (1981) in a survey of the earlier results from the negative income tax experiments. The average estimates of the compensated substitution elasticity from the experiments are uniformly lower than the average elasticities estimated in the nonexperimental literature. Since the economic efficiency costs of a particular tax or transfer plan are proportional to the compensated substitution effect, it follows that efficiency loss from a negative income tax was found to be smaller in the experiments than would have been predicted on the basis of the average elasticity estimated in nonexperimental studies.

Interestingly, the experimental estimates fall in a far narrower range than the nonexperimental estimates, though the experimental elasticities were estimated using four independent samples and a wide range of econometric models. The smaller dispersion in estimates is obvious from a comparison of the standard deviations around the mean experimental and nonexperimental point estimates. The greater robustness of the experimental estimates is presumably due to the large amount of experimentally induced random variation in net wages and nonwage income levels. This random variation reduces the effect of specification error on parameter estimates and thus minimizes the effect of using alternative econometric models. Even though the average experimental and nonexperimental elasticity estimates in table 3 are sometimes far apart, the range of experimental estimates falls well within the range observed in the nonexperimental literature. Note, for example, that the average point estimate of response in the experiments is always within one standard deviation of the corresponding point estimate from nonexperimental studies. This is, of course, primarily due to the fact that the standard deviation of nonexperimental estimates is so large. The experiments thus appear to have achieved their major goal. They have substantially reduced our uncertainty about the size of work effort reductions in response to wage rate and income changes.

Implications for Welfare Reform

The labor supply estimates reported in the previous section can be used to analyze a variety of issues about welfare reform. The most important issues concern the net budgetary costs and work effort effects of particular proposals for reform. To predict the detailed effects of a reform it is necessary to incorporate estimates from a structural labor supply model into a microsimulation model. In comparison to the large number of studies of experimental labor supply response, there have been only few studies attempting to generalize the findings from the experiments to the U.S. population. Predictions of the nationwide response to a negative income tax are rare because they are costly to obtain.

The first requirement for decent prediction is a reliable source of information about a nationally representative sample of low-income families. Most sources of data, such as Census public-use tapes or the Current Population Survey, are expensive to use. A second requirement for prediction is a computer program that can accurately define or predict both the pre-reform and post-reform situations of individuals represented in the Census file. Certain pre-reform characteristics of individuals, such as employment status, weekly hours of work, annual earnings, and unearned income, may be directly reported in the file. Other characteristics, such as taxes paid, potential welfare benefits, and marginal tax rates, must be predicted on the basis of published tax and welfare schedules and sophisticated imputation procedures. Because the United States contains 51 separate political jurisdictions with unique income tax schedules and welfare formulas, the burden of imputation is formidable. Using labor supply estimates from the negative income tax experiments (or some other source), the analyst must finally predict the amount of work effort change that will occur as a result of a reform in the transfer formula and calculate the budgetary cost of the reform, taking account of the labor supply response. Given the size of the computational burden, it is not surprising that microsimulation is seldom performed.

Table 4 shows predictions of the work effort effects and budgetary costs of four different negative income tax plans. The predictions are based on microsimulations performed by SRI International and Mathematica using estimates of work effort response from the Seattle-Denver experiment. The table shows the results of two separate simulations of response to each of the plans. The first simulation used population information covering the year 1974 and estimates of labor supply response reported in Keeley et al. (1978b). The second study used population information for 1975 and estimates of labor supply response reported in the final Seattle-Denver report (SRI International, 1983). Note that neither the baseline year nor the assumed labor supply parameters were the same in the two simulations. (Cost estimates are converted to 1985 dollars, however.) In addition, other details of the simulations differed, although the significance of these differences is difficult to interpret.[18]

The four negative income tax plans examined in the table offer two basic payment levels and two tax schedules. The lower income guarantee is 75 percent of the poverty level while the higher guarantee is one-third higher, or 100 percent of the poverty line (approximately $11,000 per year for a family of four in 1985). The plans are assumed to replace the present public assistance and food stamp programs. The lower guarantee is slightly more generous than the combined guarantees of AFDC and food stamps in a state offering the median aid benefit. However, states offer a wide range of basic aid plus food stamp payment levels, ranging from less than half to slightly more than the poverty line.[19] The two tax rates examined are 50 percent and 70 percent. By comparison, in the case of AFDC, the statutory tax rate on earnings is now 100 percent, though the statutory rate in the mid-1970s was only 67 percent. It should be stressed that effective rates have always been below statutory rates. The effective tax rate for AFDC might currently approach 70 percent, but in the mid-1970s it was as low as 30 percent (Fraker et al., 1985). The combined AFDC and food stamp effective tax rate in the mid-1970s was thus below 50 percent. In each of the negative income tax plans examined, the tax rate on unearned income is 100 percent. Positive income and payroll taxes are fully reimbursed for families with gross income below the negative income tax breakeven point. This reimbursement implies that a 100-percent-of-poverty-line guarantee assures all families of a net income equal to at least the poverty line.

Table 4
Labor Supply and Budgetary Implications of Four Negative Income Tax Plans

Negative Income Tax Plan	(1) Work Effort Change: Among Recipients	(2) Work Effort Change: In Entire Population	(3) Percent Receiving Benefits[a]	(4) Net Additional Cost[b]	(5) Population Earnings Reduction[b]	(6) (5) ÷ (4)
75% Poverty Line Guarantee/50% Tax Rate						
Husband-Wife	− 9.5%	− 1.4%	.19	$15.5	$ 9.0	.58
Female Heads	− 6.7	− 2.4	.61	.8	.4	.50
Total			.24	16.3	9.4	.58
Alternative Estimate						
Husband-Wife	− 6.5	− .8	.17	11.5	5.1	.44
Female Heads	7.9	9.0	.57	− 4.8	− 3.0	.62
Total			.22	6.7	2.1	.31
75% Poverty Line Guarantee/70% Tax Rate						
Husband-Wife	− 15.8%	− .5%	.07	$ 5.5	$ 2.2	.40
Female Heads	− 9.3	− 1.2	.51	− 1.0	.0	—
Total			.12	4.5	2.2	.49
Alternative Estimate						
Husband-Wife	− 8.0	.0	.06	1.2	− .7	—
Female Heads	5.2	11.5	.43	− 6.5	− 3.7	.57
Total			.10	− 5.3	− 4.4	.83
100% Poverty Line Guarantee/50% Tax Rate						
Husband-Wife	− 10.0%	− 3.5%	.39	$51.9	$27.1	.52
Female Heads	− 12.0	− 7.1	.73	9.2	1.8	.20
Total			.43	61.1	28.9	.47
Alternative Estimate						
Husband-Wife	− 9.8	− 3.4	.39	51.4	26.7	.52
Female Heads	− 2.2	1.5	.71	4.1	− .6	—
Total			.43	55.5	26.1	.47
100% Poverty Line Guarantee/70% Tax Rate						
Husband-Wife	− 20.6%	− 1.5%	.15	$19.6	$ 8.6	.44
Female Heads	− 14.9	− 5.3	.61	6.1	1.0	.16
Total			.20	25.7	9.6	.37
Alternative Estimate						
Husband-Wife	− 10.7	− .9	.14	14.8	5.2	.35
Female Heads	− 4.4	5.4	.57	.6	− 1.8	—
Total			.19	15.4	3.4	.22

[a]Percent of families in relevant population receiving negative income tax payments.

[b]Measured in billions of 1985 dollars. A negative sign indicates a net cost saving or net earnings increase. Estimated earnings reduction excludes the response of families who are nonrecipients before and after the reform.

Sources: Keeley et al. (1978a and 1978b). Alternative estimate from SRI International (1983).

Analysts performing the simulations assumed that the eligible population contained non-aged husband-wife families and female-headed families with children. Aged and single-person families were excluded from the simulation. The negative income tax represents a substantially different kind of reform for the three groups that would be eligible for payments. For single-parent families, the negative income tax would simply replace AFDC and food stamps, both of which are already received by a high proportion of single mothers with low incomes. For many single-parent families, the negative income tax payment might even be lower than the welfare benefit that it replaces. Two-parent families with children would be more generously treated under a negative income tax than they are under the current welfare system. These families are eligible to receive AFDC in only about half the states, and even in those states the program is less generous to two-parent families than it is to single-parent families. Childless husband-wife families would be treated far more generously under a negative income tax than they are under the current system. Such families are currently eligible to receive only food stamps and general assistance. General assistance is typically far less generous than AFDC.

The first column in table 4 shows the predicted reduction in annual hours of work among recipients of negative income tax payments. In two-parent families the work reduction under all four plans is moderately large, ranging from 6.5 percent to as much as 20.6 percent, depending on the characteristics of the plan and the details of the simulation. These estimates reflect the combined responses of both husbands and wives to the negative income tax incentives. In the first simulation there is a tendency for the percentage reduction in hours to rise strongly with increases in the guarantee and tax rate. The second simulation shows the same pattern, but it is much weaker. Note that the second simulation shows smaller work effort reductions than the first, particularly for plans with a higher marginal tax rate. In spite of their differences, both simulations show work effort reductions among husbands and wives receiving the negative income tax payments, with fairly large percentage reductions under the two plans that provide a poverty-line guarantee.

The two simulation programs do not conform in their predictions of the response among single-parent families. One simulation shows moderate to substantial hours reductions while the other shows only small reductions or even labor supply increases. It is unlikely that the inconsistencies are due to the differing labor supply parameters used.[20] They are probably caused by differences in the base year used and the assumed level of pre-reform welfare benefits. (The latter difference presumably has only a small effect in the case of husband-wife families because these families are typically ineligible for welfare benefits under the current system.) The striking differences in the predicted single-parent responses to the same negative income tax are disturbing. The differences imply that work effort estimates are sensitive to alternative techniques in simulation as well as to varying assumptions about income and substitution effects.

The second column in the table shows the predicted population response to the negative income tax plans. These predictions include work effort changes among nonrecipients as well as recipients of payments. The numbers in the column show the percentage changes in population hours of work. (Note that the percentage change in population earnings will be much smaller than the percentage hours reductions because negative income tax recipients, who account for the work reductions, have lower wage rates than nonrecipients.) Among husband-wife families, the population response is always much smaller in percentage terms than the response among recipients. The reason is obvious in view of the participation rates reported in column (3). Only a fraction of the population receives negative income tax payments, so most husband-wife families will be unaffected by welfare reform. (Neither of the simulations includes a tax increase on nonrecipients to finance the added transfer payments.) Note that the husband-wife population response rises with increases in the guarantee level but *declines* with increases in the marginal tax rate. That is, a 70 percent tax rate causes less overall work reduction than a 50 percent tax rate. The explanation for this apparently perverse result is that the participation rate in a high-tax program will be lower than in a low-tax program that has the same income guarantee. As the tax rate rises, the income cutoff point for receipt of benefits falls. Fewer families will have incomes low enough to qualify for payments, so fewer will be affected by the work disincentives implicit in the transfer formula. The estimates in the table show a conflict between the goal of providing work incentives to transfer recipients and that of providing incentives to the population as a whole.[21] Recipients can be encouraged to work through a reduction in the tax rate, but such a reduction will increase the number of recipients and hence reduce aggregate work incentives.

The trade-off between work incentives for recipients and for the population as a whole is also evident in the case of single mothers. Both simulations show that aggregate work effort is greater under a high-tax plan than under a low-tax plan with the same guarantee.[22] Both simulations also show that work effort among recipients is lower under the high-tax plan than under the low-tax plan. The two simulations do not agree, however, in predicting the sign of the overall response to a negative income tax plan among single mothers. The first simulation implies that all four negative income tax plans, including the least generous, would reduce work effort. The second implies that the plans, including even the most generous, would cause an increase in labor supply among single parents. The discrepancy is due to different assumptions about the generosity of the existing welfare system. The second simulation is based on the assumption that the current system is relatively generous, so introduction of a negative income tax would reduce benefits for a substantial fraction of current welfare recipients.[23]

The budgetary implications of the four negative income tax plans are shown in column (4). The most interesting estimates are the ones for the two plans that offer income guarantees equal to the poverty line. By definition these plans eliminate poverty among husband-wife and single-parent families. The more generous plan would cost $56 billion to $61 billion more than the current welfare and food stamp programs, or approximately 1.5 percent of GNP.[24] The less generous, high-tax plan would cost $15 billion to $26 billion more, or 0.4 to 0.6 percent of GNP. How one views these estimates depends on one's attitude toward redistribution. A person favorably inclined toward redistribution might regard the less expensive high-guarantee plan as a bargain: poverty is eliminated among families containing children, and at modest cost. Federal taxes would have to rise 2 to 4 percent to finance the plan, however, so taxpayers less favorably inclined toward redistribution would have ample grounds to oppose the reform, especially for husband-wife families.

The last two columns provide evidence that might dissuade even advocates of redistribution from suggesting a universal negative income tax. Column (5) shows the earnings reductions in response to introduction of a negative income tax. Negative values are reported in a few cases, implying that a negative income tax would actually increase aggregate earnings. But most of the entries are positive, suggesting that earnings reductions would offset at least part of the income gains to the poor produced by a negative income tax. Column (6) shows the size of the earnings change as a fraction of the net additional transfer cost of the program. The fraction is especially high in the case of two-parent families. The first simulation implies that the earnings reduction would represent 40 to 58 percent of the added transfer costs of the program for two-parent families. The second simulation implies earnings reductions ranging from 35 to 52 percent of net program costs, except in the case of the least generous program, where there is a slight earnings gain.

Husbands and wives in families receiving benefits obviously "consume" a high percentage of their benefits in the form of additional leisure or other nonmarket uses of time. While the consumption of additional leisure increases the happiness of recipient families, it simultaneously raises the cost of payments to taxpayer donors and offsets a large part of the intended redistributional impact of the payments. Even more important to some taxpayers, it raises the dependence of poor two-parent families on government transfers.

The trade-off between earnings reductions and added transfer costs is more favorable in the case of single mothers. Only the two plans with a poverty-line guarantee involve substantial added costs to taxpayers. One of the simulations shows that under these plans earnings would fall by 16 to 20 percent of additional transfer costs, while the second shows that single mothers' earnings would actually rise as a result of introduction of a poverty-line guarantee. Though I am skeptical of the second set of predictions, it seems likely that the earnings response of single mothers would be less costly to taxpayers than the response in two-parent families. This is suggested by the actual pattern of response in the Seattle-Denver experiment. During the second year of that experiment, the earnings reduction among single mothers was 39 percent of the average negative income tax payment to one-parent families, while the combined husband and wife earnings reduction was 68 percent of the average payment to two-parent families.[25] Given the same payment, the net income gain to a single-parent family would be greater than the income gain in a two-parent family. A negative income tax thus represents a more attractive reform alternative for single-parent than for two-parent families.

One of the main obstacles to improving the generosity of means-tested transfers is the knowledge that more generous benefits will reduce the earnings and self-support of the poor. The simulation results reported in table 4 suggest that this concern is reasonable for two-parent families, but is less valid in the case of single-parent families. Even though the predicted work effort reduction among husbands and wives is small, the implied reduction in earnings is a large percentage of additional transfer benefits. Using Arthur Okun's analogy, it is obvious that a negative income tax does not provide a leakproof redistributive bucket.

The bucket is nonetheless more leakproof than sometimes suggested in the nonexperimental literature. Edgar K. Browning and William R. Johnson (1984) have recently argued, for example, that the disposable money income of the top three income quintiles is depressed by $9.51 for each one-dollar increase in money income successfully transferred to the lowest two quintiles. It is depressed by this large amount because transfer recipients reduce their work effort, thus increasing the amount of money that must be transferred to raise their net incomes by one dollar. In addition, Browning and Johnson's simulations show substantial work effort reductions among taxpayers who are faced with higher tax rates as a result of the increased transfers.

Findings from the experiments suggest that the cost of redistributing one dollar to the poor must be far less than $9.51. For example, estimates in table 4 of the cost and earnings impact of the most generous negative income tax plan imply that it would cost approximately $1.89 to transfer an added dollar to the poor.[26] This estimate ignores the labor-supply response of taxpayers who must pay $1.89 in added taxes. If the net income of these taxpayers falls by $9.51, it must be the case that their net earnings fall by $7.62 (=9.51-1.89) in response to the higher tax rate. Using the assumptions of Browning and Johnson, this implies that gross earnings fall by at least $12.70.[27] The labor-supply response parameters estimated in the experiments appear inconsistent with the prediction that annual earnings of taxpayers would decline by $12.70 in response to a rise in net tax liabilities of only $1.89.

The experimental elasticity estimates reported in table 3 are in fact consistent with a slight rise in taxpayers' earnings, because the income effect of higher taxes should more than offset the substitution effect for most high-income families.[28] This is confirmed in the only microsimulation study that uses experimental labor supply parameters to predict the responses of both transfer recipients and taxpayers to the introduction of a negative income tax. In that simulation study, Betson, Greenberg, and Kasten (1982) find that the combined labor supply responses of transfer recipients and taxpayers actually cause national earnings to rise after introduction of a negative income tax. That is, the earnings gains of taxpayers more than offset the earnings reductions of transfer recipients.[29] If this conclusion is valid, the experimental results imply that the disposable money income of the top three income quintiles will fall by less than $1.89 for each one-dollar increase in money income successfully transferred to the working-age poor.[30] This estimate is, of course, far below the estimate reported by Browning and Johnson, who based their study on nonexperimental labor supply elasticities. The experimental results thus imply substantially lower costs to taxpayers of income redistribution.

Conclusions

The negative income tax plans tested in the experiments were expected to reduce work effort among participants, and they did so. The work reductions were probably smaller than most opponents of a negative income tax had feared, but larger than advocates had hoped. In comparison to predictions of work effort response based on prior nonexperimental research, the actual response to the tested plans was small. But the response was negative even among women previously receiving public welfare, with all of its attendant work disincentives. The estimates of income and substitution elasticities obtained in the experiments fall well within the very broad range of estimates obtained in nonexperimental studies. Moreover, the experimental estimates appear to be far more robust. That is, they fall within a narrow range even when estimated using different samples and alternative econometric models. With the exception of the income elasticity estimated for women, the average experimental elasticities are lower in absolute value than corresponding nonexperimental estimates. In particular, the compensated substitution elasticity is only a fraction of the average elasticity estimated in nonexperimental studies, implying that the efficiency losses for redistribution to the able-bodied poor would be lower than could be predicted from the average nonexperimental estimates of response.

It has been argued, by Anderson (1978) and Murray (1984) among others, that the findings of the experiments greatly understate the long-run response to a negative income tax. While there is some evidence from the experiments themselves that the long-run impact is indeed understated, the evidence is neither as strong nor as unambiguous as these critics argue. The permanent income effect of negative income tax payments was almost certainly underestimated in the experiments, but the substitution effect of the tax rates was probably overstated, at least among wives. While it is true that participants in the experiments may not have had time to fully adjust their labor supply to its long-run equilibrium value, it is equally true that the experiments did not observe the long-run response of employers to a smaller supply of low-wage labor.[31] Moreover, at least part of the apparent labor supply response in the experiments is known to have been a reporting phenomenon rather than a true reduction in work effort. (It is arguable whether the protections against income misreporting in a national program would be greater or less than those available in an experiment.) Given these potentially offsetting biases, the long-run impact of a modest negative income tax is probably understated by no more than one-third by simple extrapolation of the experimental results.

The estimates obtained in the experiments have a number of implications for reform of the welfare system, especially reform that raises the generosity of benefits. The findings suggest that benefit increases would cause only moderate reductions in aggregate hours of work and even smaller reductions in aggregate earnings. But even if the overall work reduction is small, the resulting earnings loss among recipient breadwinners would represent a large fraction of the higher payments sent out to low-income families. Earnings reductions would therefore offset a substantial part of the income gain from more generous transfers.

The arithmetic of reform is especially melancholy in the case of two-parent families, where earnings reductions might represent 50 to 60 percent of the added cost of new transfers. A simple and moderately generous negative income tax appears to be far more feasible for single-parent families. The earnings response of single mothers is small or even slightly positive. The experiments thus provide some support for offering generous benefits only to families whose earnings are less responsive to work disincentives. George Akerlof (1978) has argued, for example, that high-benefit, low-tax transfer formulas should be made available to only the least responsive families so that benefits can be more generously provided to those in greatest need. The results of the experiments support Akerlof's argument that the trade-off between higher benefits and lower work effort would be less painful under a system of separate transfer formulas for one- and two-parent families. (As Akerlof also points out, this is a fair description of the current welfare system.) Unfortunately, such a system provides clear incentives for families to change their composition in order to become eligible for the more generous transfer formula.

The findings from the experiments also point up a conflict between creating work incentives for transfer recipients and for the population as a whole. If a major goal of a transfer formula is to provide work incentives for recipients, the findings imply that relatively low tax rates are desirable. If the goal is to reduce disincentives for the entire population, a much higher tax rate is preferable because it minimizes the size of the population subject to work disincentives. This trade-off is clearest in the case of husband-wife families, where reductions in the marginal tax rate (given a fixed and plausible guarantee level) cause rapid increases in the population eligible to receive benefits. For single-parent families the trade-off is less clear since so many single mothers are eligible to receive payments, even at low guarantee levels. Hence, reductions in the marginal tax rate do not cause such rapid increases in the proportion of one-parent families eligible to receive payments.

If the experiments have inspired pessimism about our ability to reduce poverty through a system of pure cash transfers, they have also stimulated an examination of alternatives to a negative income tax. One way to minimize the adverse earnings effects of generous transfers is to require recipients to work. The Carter administration proposed to do this through a program of guaranteed public sector jobs for welfare recipients who were expected to work. Recipients refusing to work would have been denied benefits under the more generous transfer formula and forced to rely on benefits computed under a less generous formula. While the Carter proposal would have reduced or even eliminated the adverse earnings impact of more generous transfers, it would have involved substantial additional costs in order to finance the guaranteed jobs program. Some of these costs can be avoided under workfare, which essentially requires welfare recipients to work but does not pay them anything in addition to their current welfare grant if they do so. Recipients who decline work can have their grants reduced or eliminated. The negative income tax experiments obviously shed little if any light on the effects of this kind of work requirement.

Wage subsidies and earnings subsidies represent an alternative approach to redistribution. A worker eligible for a wage subsidy receives a transfer payment that grows rather than declines as hours of work rise. Not only does the program redistribute income to the poor, but it offers larger transfers to breadwinners who work longer hours. The labor supply response to wage subsidies is thus assumed to reinforce rather than offset the direct redistributive effect of the transfer payments. The response estimates obtained in the negative income tax experiments can be used to predict the effects of wage subsidy plans as well as negative income tax plans. The elasticity estimates reported in table 3 do not appear especially encouraging for a wage or earnings subsidy scheme. The labor supply functions estimated in the experiments are vertical or backward-bending. Much of the response to negative income tax payments was caused by a reliably estimated income effect. Any wage or earnings subsidy thus has the potential to encourage work reductions among those breadwinners who would receive the largest subsidies, that is, those now working the longest hours. In a simulation study of the impact of wage-rate subsidy schemes based on labor supply estimates from the Seattle-Denver experiment, analysts have found that subsidy plans actually reduce hours and earnings in recipient families (Betson and Bishop, 1982). Contrary to the expectations of subsidy advocates, the work response to wage subsidies—like the response to negative income tax payments—tends to offset the direct redistributive impact of the transfers.

The experiments have confirmed that good deeds are not costless. Income redistribution to the poor has an efficiency price. The price is far lower than pessimists predicted, but it certainly exceeds zero. The reaction of policymakers and policy analysts to this set of findings is interesting. They seem far more impressed by our certainty that the efficiency price of redistribution is positive than they are by the equally persuasive evidence that the price is small.

[1]The results in table 2 were not based on an identical statistical specification across experiments, nor were the estimation samples selected with identical criteria. The estimates reflect the responses to negative income tax plans in the middle two years for the New Jersey experiment, in the entire three years of the Rural and Gary experiments, and in the second (or middle) year of the Seattle-Denver experiment. For estimates based on a similar model and set of sample selection criteria, see Robins (1985), who reports very similar results. I slightly prefer the results reported here because they reflect the judgments of analysts who were most familiar with data from the individual experiments.

[2]See Moffitt (1979, p. 479) and Robins and West (1980b, pp. 23, 25, and 27).

[3]SRI International (1983) p. 177.

[4]The after-tax income of eligible families was raised by somewhat more than $900. The estimated reduction in gross earnings is $1,800 but the implied reduction in net earnings is probably 10 to 20 percent below that figure.

[5]Greenberg, Moffitt, and Friedman (1981, p. 586).

[6]Greenberg and Halsey (1983, pp. 400-05). In an unpublished analysis of underreporting based upon earnings records from the Social Security Administration rather than the unemployment insurance system, SRI obtained similar results. Underreporting of income to the experiment caused a very slight overstatement of the true earnings reduction among husbands and single mothers and a somewhat larger overstatement of the reduction among wives in two-parent households. Some observers argue that the experiments' experience with income misreporting is not relevant in a fully operational national program, since a national program would have access to employer-reported earnings information, such as that available to the Social Security Administration. While it is possible to use Social Security and unemployment insurance administrative records to verify the earnings reductions estimated from interview data, it would be impractical to rely on these same administrative records to compute monthly negative income tax payments. The Social Security Administration and state unemployment agencies obtain individual earnings records only with a lag, which can range up to 18 months. This is clearly too long to permit the timely calculation of negative income tax benefits. Hence, any practical system of monthly (or bimonthly) transfer payments must rely on self-reported earnings information, at least to some degree. For that reason, the experimental findings on income underreporting are applicable to a wide range of feasible welfare reform plans.

[7]A very small number of families was enrolled for 20 years, but this sample is probably too small to yield useful results.

[8]Committee on Ways and Means (1986), pp. 373-74.

[9]Office of Income Security Policy (1983), p. 6.

[10]To illustrate the problem, consider the earnings reduction among Seattle-Denver husbands during the third experimental year. Men assigned to the lowest guarantee/50% tax plan reduced their earnings by an average of $962, while men assigned to the highest guarantee/50% tax plan reduced their earnings by only $592. An explanation for this perplexing pattern of response is provided by the sample assignment plan. Whereas 96% of men in the less generous plan had preexperimental income below $7,000, only 26% of men in the more generous plan had income below that level. If we estimate the effect of both plans separately for each preexperimental income level, we can compute what the expected responses would be in two samples with an identical income distribution. Suppose we consider a sample that has the income distribution of the combined samples assigned to the two negative income tax plans just mentioned. The expected response to the low guarantee/50% tax plan is an increase in earnings equal to $753 per year, while the expected response to the high guarantee/50% tax plan is an earnings reduction of $1,994. Both predictions are extremely imprecise because of the small number of men within particular income classes assigned to one or another of the plans. Clearly, the sampling plan had an enormous impact on the pattern and precision of estimated responses to the two plans.

[11]Results from this procedure are statistically imprecise because there are only a few observations in each cell when all conceivable interaction effects are estimated.

[12]A worker typically faces a segmented linear rather than a strictly linear budget constraint defining the trade-off between leisure and consumption. Each linear segment is defined by a slope (equal to the net or after-tax wage rate) and an intercept term referred to as "virtual income." If a worker faced a strictly linear budget constraint, the intercept would be equivalent to the amount of nonwage income to which the worker is entitled at zero hours of work.

[13]By definition, Δw and ΔY are directly correlated with preexperimental work effort since they are defined at the preexperimental level of hours. The correlation is nonetheless smaller than it would be in nonexperimental data.

[14]Strictly speaking, it would not concern us if Δw, ΔY, w, and N are correlated with observed variables so long as those variables are included in the estimation equation. As a practical matter, however, a high correlation between, say, w and X makes it difficult to estimate precisely the separate effects of w and X on H.

[15]The studies are Keeley et al. (1978b), Keeley and Robins (1980), Robins and West (1980a), Burtless and Greenberg (1982), Johnson and Pencavel (1982), SRI International (1983), and Johnson and Pencavel (1984). Labor supply elasticities for most of these studies are reported in Keeley (1981), pp. 159-67.

[16]The 21 estimates were not obtained in 21 different studies. Several studies reported separate labor supply estimates for different racial groups. For example, both New Jersey and Rural experimental studies often reported separate labor supply parameters for different racial groups.

[17]I include all elasticities reported by Killingsworth (1983) on pp. 119-122 and pp. 193-197 from U.S. studies where it is possible to compute them. Some individual studies provide several estimates of labor supply response; each response estimate is included with equal weight.

[18]For example, the first simulation considered the response of household heads aged 18 to 58, while the second considered responses of household heads between 16 and 65. The second simulation also used a significantly different method of imputing transfer benefits, which had important consequences for defining the pre-reform situation of low-income families (see below). Standard errors of the simulated national labor supply responses are reported in SRI International (1983), p. 181.

[19]See Committee on Ways and Means, (1986), pp. 370-75.

[20]The labor supply elasticities assumed in the two simulations do not differ very much for female heads. The uncompensated and compensated substitution elasticities and the total income elasticity were −0.03, 0.13, and −0.15, respectively, in the first simulation; they were −0.04, 0.17, and −0.22 in the second. For husbands, in the first simulation the elasticities were 0.02, 0.10, and −0.08; in the second they were −0.13, 0.09, and −0.22. For wives, in the first simulation the elasticities were 0.00, 0.22, and −0.22; in the second they were −0.11, 0.20, and −0.31.

[21]See also Levy (1979) and Moffitt (1985) for a discussion of this issue.

[22]This corresponds to Levy's (1979) findings with respect to AFDC but contradicts Moffitt's (1985) simulation of the effect of a pure negative income tax using nonexperimental labor supply elasticities.

[23]In fact, a majority of current single-parent welfare recipients is predicted to be worse off under three of the four plans examined. Even the most generous plan—offering a 100-percent-of-poverty-line guarantee and 50 percent tax rate—is predicted to make more than one-third of current welfare recipients worse off. See SRI International (1983), p. 189.

[24]These statements may understate the cost of a poverty-line income guarantee in the mid 1980s. The simulations are based on population responses in the mid 1970s when the employment rate of married and single mothers was somewhat lower. Since the labor supply response of women accounts for an important share of the net cost of a more generous program, the budgetary impact of a negative income tax could be higher in the 1980s.

[25]SRI International (1983), pp. 117 and 144.

[26]Table 4 contains four estimates of the net cost of guaranteeing a poverty-line income, two based on an assumed tax rate of 50 percent and two based on a tax rate of 70 percent. Column 6 shows the ratio of earnings reductions to net additional budget outlays. The highest reported ratio for the poverty-line plans is 0.47. This implies that \$1.00 in additional transfer benefits causes a \$0.47 reduction in earnings, suggesting that net income is only \$0.53 (or \$1.00 −\$0.47) higher than it would be without the additional transfers. By implication, taxpayers must spend \$1.89 to raise the net incomes of the poor by \$1.00.

[27]The marginal tax rate in the top three income quintiles is estimated to be about 40 percent (see Browning and Johnson, 1984, p. 184). With this tax rate, a \$12.70 decline in gross wages yields a \$7.62 decline in net wages.

[28]Oddly, Browning and Johnson argue that their simulation predictions are consistent with labor supply elasticities estimated in the experiments (Browning and Johnson, 1984, pp. 190-91). In fact, Browning and Johnson's assumed labor supply elasticities (p. 188) differ markedly from the experimental elasticities reported in table 3. The discrepancies are especially notable in the lowest income quintile.

[29]The results of this simulation are described in Betson, Greenberg, and Kasten (1982), p. 200. For a related discussion, see Betson and Greenberg (1986). We should be cautious in accepting simulations of the taxpayer response to tax increases that are based on response parameters obtained in the negative income tax experiments. The experiments enrolled low-income families; most taxes are paid by middle- and high-income families.

[30]We should carefully distinguish between the earnings effects of a tax increase and the welfare or economic efficiency effects. Even though the gross earnings of taxpayers might rise as a result of a tax increase, the welfare of such taxpayers must decline by at least as much as the added revenue raised by the tax. Thus, even if the net income of taxpayers falls by less than $1.89, the welfare of taxpayers must fall by more than $1.89. Depending on the size of the compensated substitution effect and existing marginal tax rate, the welfare loss could substantially exceed $1.89.

[31]In the long run, for example, wage offers by employers might be higher or the unemployment rate among nonrecipients of a negative income tax might be lower. The latter effect would occur if negative income tax recipients and nonrecipients are in competition for a limited number of jobs.

References

Akerlof, George A. "The Economics of 'Tagging'," *The American Economic Review,* 68 (March 1978), pp. 8-19.

Anderson, Martin. *Welfare,* Stanford, CA: Hoover Institution Press, 1978.

Ashenfelter, Orley, and James Heckman. "The Estimation of Income and Substitution Effects in a Model of Family Labor Supply," *Econometrica,* 42 (January 1974), pp. 73-85.

Betson, David M., and John H. Bishop. "Wage Incentive and Distributional Effects," in R.H. Haveman and J.L. Palmer, eds., *Jobs for Disadvantaged Workers: The Economics of Employment Subsidies,* Washington, DC: The Brookings Institution, 1982, pp. 187-209.

Betson, David M., and David Greenberg. "Labor Supply and Tax Rates: Comment," *The American Economic Review,* 76 (June 1986), pp. 551-56.

Betson, David M., David Greenberg, and Richard Kasten. "A Simulation Analysis of the Efficiency and Distributional Effects of Alternative Program Structures: The NIT Versus the CIT," in I. Garfinkel, ed., *Income-Tested Transfer Programs: The Case For and Against,* New York: Academic Press, 1982, pp. 175-203.

Borjas, George, and James Heckman. "Labor Supply Elasticities and Public Policy," *Proceedings of the 31st Annual Meeting of the Industrial Relations Research Association,* 1978.

Browning, Edgar K., and William R. Johnson. "The Trade-Off Between Equality and Efficiency," *Journal of Political Economy,* 92 (April 1984), pp. 175-203.

Burtless, Gary, and David Greenberg. "Inferences Concerning Labor Supply Behavior Based on Limited-Duration Experiments," *The American Economic Review,* 72 (June 1982), pp. 488-97.

Burtless, Gary, and Jerry A. Hausman. "The Effect of Taxation on Labor Supply: Evaluating the Gary NIT Experiment," *Journal of Political Economy,* 86 (December 1978), pp. 1103-30.

Committee on Finance, U.S. Senate. *Welfare Research and Experimentation.* Hearings of November 15-17, 1978. Washington, DC: U.S. Government Printing Office, 1978.

Committee on Ways and Means, U.S. House of Representatives. *Background Material and Data on Programs within the Jurisdiction of the Committee on Ways and Means: 1986 Edition,* Washington, DC: U.S. Government Printing Office, 1986.

Fraker, Thomas, Robert Moffitt, and Douglas Wolf. "Effective Tax Rates and Guarantees in the AFDC Program, 1967-82," *The Journal of Human Resources,* 20 (Spring 1985), pp. 264-77.

Greenberg, David, and Harlan Halsey. "Systematic Misreporting and Effects of Income Maintenance Experiments on Work Effort: Evidence from the Seattle-Denver Experiment," *Journal of Labor Economics,* 1 (October 1983), pp. 380-407.

Greenberg, David, Robert Moffitt, and John Friedmann. "Underreporting and Experimental Effects on Work Effort: Evidence from the Gary Income Maintenance Experiment," *The Review of Economics and Statistics,* 63 (November 1981), pp. 581-89.

Johnson, Terry R. and John H. Pencavel. "Forecasting the Effects of a NIT Program," *Industrial and Labor Relations Review,* 35 (January 1982), pp. 221-234.

________."Dynamic Hours of Work Functions for Husbands, Wives, and Single Females," *Econometrica,* 52 (March 1984), pp. 363-89.

Keeley, Michael C. *Labor Supply and Public Policy: A Critical Review,* New York: Academic Press, 1981.

Keeley, Michael C., and Philip K. Robins. "The Design of Social Experiments: A Critique of the Conlisk-Watts Assignment Model," in R. Ehrenberg, ed., *Research in Labor Economics, Vol. 2,* Greenwich, CT: JAI Press, 1980.

Keeley, Michael C., Philip K. Robins, Robert G. Spiegelman, and Richard W. West. "The Labor Supply Effects and Costs of Alternative NIT Programs," *The Journal of Human Resources,* 13 (Winter 1978a), pp. 3-36.

________. "The Estimation of Labor Supply Models Using Experimental Data," *The American Economic Review,* 68 (December 1978b), pp. 873-87.

Kehrer, Kenneth C. *The Gary Income Maintenance Experiment: Summary of Initial Findings,* (mimeo) Gary, IN: Indiana University, 1977.

Killingsworth, Mark R. *Labor Supply,* New York: Cambridge University Press, 1983.

Levy, Frank. "The Labor Supply of Female Heads, or AFDC Work Incentives Don't Work Too Well," *The Journal of Human Resources,* 14 (Winter 1979), pp. 76-97.

Moffitt, Robert A. "The Labor Supply Response in the Gary Experiment," *The Journal of Human Resources,* 14 (Fall 1979), pp. 477-87.

________. "A Problem with the Negative Income Tax," *Economics Letters,* 17 (1985), pp. 261-65.

Moffitt, Robert A., and Kenneth C. Kehrer. "The Effect of Tax and Transfer Programs on Labor Supply: The Evidence from the Income Maintenance Experiments," in R.G. Ehrenberg, ed., *Research in Labor Economics, Vol. 4,* Greenwich, CT: JAI Press, 1981, pp. 103-50.

Murray, Charles. *Losing Ground: American Social Policy 1950-1980,* New York: Basic Books, 1984.

Office of Income Security Policy, U.S. Department of Health and Human Services. *Overview of the Seattle-Denver Income Maintenance Experiment Final Report,* Washington, DC: U.S. Government Printing Office, 1983.

Rees, Albert. "An Overview of the Labor-Supply Results," *The Journal of Human Resources,* 9 (Spring 1974), pp. 158-180.

Robins, Philip K. "A Comparison of the Labor Supply Findings from the Four NIT Experiments," *The Journal of Human Resources,* 20 (Fall 1985), pp. 567-582.

Robins, Philip K., and Richard W. West. "Labor Supply Response to the Seattle and Denver Income Maintenance Experiments: Alternative Estimates of a Structural Model," (mimeo) Menlo Park, CA: SRI International, January 1980a.

________. "Labor Supply Response to the Seattle and Denver Income Maintenance Experiments: Analysis of Six Years of Data," (mimeo) Menlo Park, CA: SRI International, May 1980b.

SRI International. *Final Report of the Seattle-Denver Income Maintenance Experiment, Volume 1: Design and Results,* Washington, DC: U.S. Government Printing Office, 1983.

U.S. Department of Health, Education, and Welfare. *Summary Report: Rural Income Maintenance Experiment,* Washington, DC: U.S. Government Printing Office, 1976.

Discussion

*Orley C. Ashenfelter**

Having been commissioned to write a paper similar to the one by Gary Burtless at a much earlier state in the development of the negative income tax experiments (Ashenfelter 1978), I had hoped to see a major effort to address some of the puzzles that were evident to any serious scientist examining the early results of those experiments. I am afraid that Burtless has passed over all of these basic issues in his apparent determination to reach strong and definite conclusions about public policy. The result is that Burtless's paper is at best an incomplete catalogue of the research that has already been done with the negative income tax experiment data. At worst it leaves the impression that many of the important reasons for experimentation have now disappeared. Quite to the contrary, I believe most of the important research with social experiments of this type remains to be done. Careful analysis of the data already available and the design and implementation of new and better experiments could have enormous payoffs for our understanding of the effects of public policies on the poor *and* on our understanding of behavior in the labor market.

In order to demonstrate the veracity of my assertion in the limited space available, I will simply take up the two most important issues that troubled me in my review of the rural negative income tax experiment a decade ago. These issues are, as it turns out, of fundamental importance for the interpretation of the results of a negative income tax experiment, and they seem to remain as unresolved now as they were a decade ago.

First, what is the size of the effect of a negative income tax on hours worked? Burtless produces a handy table 2 that, at first blush, provides

*Professor of Economics and Director of the Industrial Relations Section, Princeton University.

the answer to this question for the programs detailed in table 1. Unfortunately, the data in table 2 are taken from the statements by program participants to the survey research houses responsible for data collection in these experiments. A key point about a negative income tax program, however, is that, like a positive income tax, it sets up an incentive for workers to underreport their incomes. The more they can reduce income reported to the experimenter, the greater will be their transfer payments.

In the reports of the New Jersey and rural negative income tax experiments, underreporting was little discussed. As Burtless states, there is some research on this issue in the Gary and Seattle-Denver experiments that indicates that income underreporting is a major (and perhaps the only) cause of the observed decline in earnings in both of these experiments. Of course, the design of these experiments did not incorporate the likelihood that income underreporting would be a serious problem, so the way it is studied is indirect. In particular, earnings from government administrative records are used to measure "true" earnings and then these are compared against the survey data.

The conclusion that Burtless draws from his appraisal of the studies of underreporting is that, "even if misreporting bias causes an exaggeration of the efficiency loss from a negative income tax . . . an earnings reduction caused by underreporting is just as costly to taxpayers as a reduction caused by a genuine reduction in work effort." Although Burtless is only adopting the same conclusion as many others, it seems to me to be in serious error. After all, a genuine negative income tax program *will* operate from government administrative reports on income. Thus, payments in a genuine negative income tax program would be based on the "true" records used by the experimenter here to establish the extent of underreporting. Unless participants actually did change their labor supply behavior or found a way to misreport their income to government officials, it is possible that the additional program costs of a genuine negative income tax scheme attributable to reductions in work effort might be very small. Who is to say whether there would be any labor supply response, further income underreporting, or neither, if an experiment with conventional administrative procedures were implemented? Only an experiment fully informed at the design stage about the possibility for income underreporting, and that tested for its effect, would shed any light on this critical issue. Sadly, the design of none of these experiments was so informed.

A second important issue revolves around the determination of precisely why a labor supply response is produced by a negative income tax experiment. To economists there are effects associated with (a) the size of the tax rate in the program and (b) the generosity of the program. Sorting out these effects is an issue of high priority if the results of the experiments are to be used to predict the expected response to a program not yet tested. No less importantly, for an economist, the incentive effects of a negative income tax program must operate through variations in the tax rate and generosity of the program, if we are to put much faith in the conventional models of labor supply often used to analyze these issues.

The reports of the results of all these experiments rarely, if ever, provide simple, nonparametric two-way contrasts of labor supply behavior by experimentals and controls. Most analysts estimate parametric models (of the form (4) in Burtless's paper) before providing any tabulation of nonparametric results. A partial exception is the final report on the Seattle-Denver income maintenance experiments (1983). A key finding there is that simple two-way contrasts show no clear evidence that higher tax rates are associated with higher labor supply responses than lower tax rates. Furthermore, no clear relationship was found between program generosity and labor supply responses. My guess is that, at best, the reported magnitudes of income and substitution effects in Burtless's table 3 are based on parametric models so weakly related to the data available that most of the results mainly reflect prior views of the experimenter, and not the actual data. At the very least we are owed some notion of the extent to which the data discipline these results, rather than the prior views of those who calculated them. I find it quite surprising that, a decade after this research was begun, it is still difficult to find out precisely what it is that nonparametric models fit to the basic experimental data reveal, if anything, about the nature of income and substitution effects on labor supply.

References

Ashenfelter, Orley C. "The Labor Supply Response of Wage Earners," in John L. Palmer and Joseph A. Pechman, eds., *Welfare in Rural Areas: The North Carolina-Iowa Income Maintenance Experiment,* Washington, DC: The Brookings Institution, 1978, pp. 109-48.

SRI International. *Final Report of the Seattle-Denver Income Maintenance Experiment, Volume 1: Design and Results,* Washington, DC: U.S. Government Printing Office, 1983.

Discussion

*Robert E. Hall**

No topic could be further removed from discussion in Washington today than a guaranteed income for all Americans, financed by a steep tax on the first few thousand dollars of income. Instead, the whole thrust of policy has been toward tightly limited categorical benefits financed by low marginal rates on all earnings. Hence, the experiments discussed by Burtless and the other authors at this conference cannot be seen now as bearing on policy choices. Rather, they provide data points for scientific investigations of the responses of families to changing economic incentives. I agree strongly with the basic theme of the Burtless paper that the main focus of research should be the incorporation of experimental data into structural labor supply estimation, and not the evaluation of the effects of the particular plans that were the subjects of the experiments.

Burtless in his paper notes the bias toward a finding of high elasticities of labor supply in the experimental data because it was in the interest of the subjects to understate their earnings in order to enlarge their payments. He reviews the attempts that have been made to measure the bias by measuring earnings from extrinsic data. In some cases, such as the Gary experiment, most of the observed decline in hours of work appears to be underreporting. In the Seattle-Denver experiment, primary earners did not underreport but secondary earners did. The reader is left with some unresolved questions: Why go on to use the data that are contaminated by known underreporting later in the paper? Why is underreporting rampant in some instances yet absent in others, where the incentive is just as strong? As the paper stands, it ap-

*Professor of Economics and Senior Fellow, Hoover Institution, Stanford University.

pears that labor supply responses should be studied directly with the extrinsic data, ignoring the reports of the subjects themselves, or at least that studies should be confined to those cases where the problem of underreporting is known to be mild.

The experimental data dramatically improve the variation in the right-hand variables in labor supply estimation. Moreover, thanks to random assignment, the variation is fully exogenous. Hence, both the bias and the randomness of estimated labor supply elasticities are smaller with experimental data than with survey data. In this respect, the scientific value of the experiments has been enormous.

Before looking at the labor supply findings, Burtless considers the biases that arise from the temporary nature of the experiments. He notes that adjustment costs and temporary income subsidies cause the experimental data to understate the long-run effect on labor supply, but that intertemporal substitution causes the data to overstate the long-run effect. His conclusion is that the net effect is an understatement of the response, but I see this as an unsettled issue.

For men, Burtless observes that econometric work has almost universally found that both the substitution and income responses of labor supply are substantially smaller in the experimental data than in survey data. That observation confirms the misgivings that veterans of labor supply estimation in survey data have always had—wages and preferences favoring work are positively correlated in the population. The cross-sectional labor supply function has a positive wage elasticity even if the labor supply function of each individual has zero elasticity. The comparison of two men, one earning \$10 per hour and the other \$5, shows the former working more hours than the latter. Conclusions about the labor supply functions of either of the men are hard to reach. On the other hand, in the experimental data, we can study a man earning \$10 per hour before the experiment, who starts paying a 50 percent tax and hence faces a decline in his wage to \$5 per hour. His decline in hours of work is unambiguously a measure of his labor supply elasticity.

For women, the results collected by Burtless show much smaller substitution responses in the experimental data than in the survey data, by an order of magnitude. The high substitution elasticities found in survey data for women are apparently the result of an even higher correlation between wages and preferences favoring work than is the case for men. However, the income responses in the experimental data are larger than those found in survey data, the opposite of what is found for men.

Burtless goes on to apply the labor supply findings to evaluate the effects of possible negative income tax programs for the U.S. economy. As I mentioned at the outset, this exercise is of relatively minor importance, since no plan of this type has any chance of active consideration, but still it is an interesting way to draw out the implications of the labor supply findings. One of the interesting things we learn as part of the exercise is that the biggest uncertainty about the effect of a move to a negative income tax as a replacement for state-administered welfare programs is the economic characterization of those programs, not the elasticities of labor supply. From the point of view of his table 4, it is just as important to carry out research on benefit levels and implicit tax rates for the existing systems in 51 states as it is to process data from the experiments.

In table 4 and earlier in the paper, Burtless invites evaluation of negative income tax programs in terms of the ratio of earnings reductions to "costs." I find this type of calculation a mystery. A negative income tax is a lump-sum benefit (a demo-grant) paid to every family, financed in part by a tax at a high rate on the first few thousand dollars of earnings of all workers and in part by the general tax system. The cost in terms of resources—government purchases of goods and services—is zero. We could also talk about the cost in the sense of the deadweight burden of the tax, but this is not what Burtless does. Yet another sense of the cost would be the total amount of the lump-sum benefits paid to all families. Again, this is not what he considers. Rather, he makes an economically arbitrary distinction between the revenue raised by the new tax on earnings and the revenue from the existing tax. The "cost" is the difference between the lump-sum benefits and the revenue from the new part of the tax. I cannot see any economic sense in which this is a cost.

Burtless seeks some kind of normalization of the aggregate earnings reduction so that it can be expressed as a percent rather than a total dollar amount. However, his choice of normalization, the "cost," is small, because most negative income tax plans generate most of the revenue needed to finance their lump-sum benefits from their own taxes. Hence his normalized earnings effects are very large. A much superior normalization, in my view, is simply the total amount of earnings. In other words, the percent reduction in earnings is the best normalized way to express the magnitude of the earnings reduction.

An important finding of Burtless's study and many earlier ones is that there is a positive relation between the tax rate and total work effort, even though each worker's labor supply function has a negative relation between his tax rate and his work effort. The reason is that a higher tax rate means that a smaller fraction of workers are subject to the tax. This finding was the explicit rationalization for welfare changes introduced in the early 1980s, when implicit tax rates for the welfare system were raised dramatically.

Burtless notes but does not stress the cruel dilemma of income supplements—under a straight negative income tax, most of the benefits go to two-parent families, yet correcting this inequality strongly subsidizes the splitting up of families. Since the conference failed to resolve the central question of the impact of welfare and negative income tax incentives on family splitting, it is hard to know how to balance the two goals of helping the neediest most and providing incentives for intact families.

In this paper, Burtless has done a commendable job in bringing together the results of a huge body of research and reducing it to its essential elements.

The Income Maintenance Experiments and the Issues of Marital Stability and Family Composition

*Glen G. Cain**

Between 1968 and 1978 four negative income tax experiments were conducted; they were designed to measure labor supply and earnings. The experiments were not designed to measure the effects of government programs on such demographic behavior as marital dissolution, fertility, family composition, or the decision to marry or remarry. Nevertheless, the data from the experiments have been used to analyze all these family issues, and they are the subject of this paper.

The essential reform examined in the negative income tax experiments was the extension of a guaranteed minimum income to poor families with an able-bodied, non-aged husband or father as the potential provider. The income plans tested in the experiments were expected to lead to reductions in the labor supply and earnings of the participating married-couple families. By a twist of fate, however, the most influential research finding of the experiments turned out to be not about labor supply but about marital stability, a family issue. The findings on labor supply showed reductions neither large enough nor small enough to permit a definitive verdict about the negative income tax. In contrast, the findings about marital stability appeared decisive.

The most important research on marital stability was conducted by Groeneveld, Hannan, and Tuma, based on the Seattle-Denver income maintenance experiment.[1] They concluded that the negative income tax increased marital dissolutions, even though it had been designed to

---

*Professor of Economics, University of Wisconsin. For help in getting the data the author is grateful to Katherine Dickinson, Mario Lopez-Gomez, Philip Robins, Daniel Weinberg, and Richard West. James Albrecht, Edward M. Gramlich, and Nancy B. Tuma provided helpful comments. Douglas Wissoker provided excellent counsel as well as programming and statistical help.

cover and assist families headed by married couples as well as families headed by women. Indeed, their finding applied to a negative income tax plan of the same level of generosity as the prevailing aid to families with dependent children (AFDC) plan. This conclusion was unambiguously unfavorable to advocates of a negative income tax that would cover married couples, for two important reasons. First, increased marital breakups among the poor would increase the numbers on welfare and the amount of transfer payments, principally because the separated wife and children would receive higher transfer payments. Second, marital dissolutions and the usual accompanying absence of fathers from households with children are generally considered unfavorable outcomes regardless of whether or not the welfare rolls increase.

Besides appearing decisive, the experimental findings about marital stability were dramatic: the reported increase in marital splits was large; it was counter to the outcome hoped for and expected by advocates of the negative income tax reform; and it was counter to the predictions of social scientists, and in particular economists. The dramatic findings received considerable attention by the press and intense scrutiny by scholars, who were skeptical but eventually accepted the findings.

This review of the negative income tax experiments offers two main messages. The first is that the evidence about the issue of marital stability is not decisive, or even persuasive. A second message is that family issues such as marital stability are not well suited to experimental research. The costs of a properly designed experiment seem too high.

Social Experimentation and Family Issues: The Case of Marital Stability

The belief that marital stability among low-income families has been adversely affected by our current welfare system seems firmly entrenched, even though empirical evidence in support of this belief has been difficult to marshal.[2] The general upward trend in divorce, separation, and female headship of families throughout this century[3] applies to all income strata; welfare programs are likely to be a factor in marital stability only among the lower half of the family income distribution and only during the last 25 years or so. In recent decades the generosity of welfare programs systematically, although not steadily, increased in ways that tended to lower the financial cost of marital dissolution to a married couple with children. For a mother, the income from welfare, which may include such in-kind payments as food stamps, Medicaid, and housing subsidies, as well as the cash payments from AFDC, provides an alternative to her husband's income. Welfare is likely in poor families to exceed the income the wife receives from the husband.

Moreover, mothers in poor families are less likely to be capable of earning enough to be self-supporting at a level of income that exceeds welfare. Finally, the availability of welfare essentially required the husband-father to leave the marriage. For a departing father, the welfare programs provided a de facto if not a legal alternative to alimony and child support.

The increase in marital dissolutions is, however, only one of the trends contributing to the increase in the proportion of families headed by a woman. Increases in the number of unwed mothers, in the length of time that mothers remain without husbands (regardless of whether they are divorced or never married), and in the proportion of mothers who establish separate households, are all sources of increased female headship and of welfare recipiency.

In summary, the spread and increased generosity of welfare programs have reduced the price (or cost) of marital dissolution. The AFDC program has been decidedly nonneutral regarding marital dissolutions in its dispensing of transfer payments and other benefits.[4] One advantage widely claimed for a negative income tax was that it would move the income maintenance system toward neutrality in marital decisions.

An Economic Framework for Analyzing Marital Stability and the Negative Income Tax Experiments

To determine how marital stability among the population of already married couples will be affected by a negative income tax, two regimes must be compared:

- The current system of welfare programs, referred to as AFDC, which provides a net subsidy to a dissolved marriage, given the presence of dependent children and assuming that the divorced woman meets the income criteria for eligibility.
- A negative income tax regime, in which the current system is amended to add welfare assistance for married couples who meet the income criteria for eligibility.

We may assume for the negative income tax, as we did for AFDC, that the only way it affects marital stability is by the income changes that it brings about. Income changes in turn induce an "income effect" in a regime where the income receipts are neutral with respect to marital status, and a "price effect" that refers to the nonneutrality of the income change with respect to marital status. Even with the simplifying assumption that income changes are all that matter, three differences between the negative income tax and AFDC are critical.

First, a negative income tax provides transfer payments for married couples whose incomes are low. Economists and sociologists appear to agree that modest increases in the incomes of married couples ought to have a positive effect on marital stability. They find both an empirical negative relation between family income and the probability of a marital dissolution and a theoretical argument that poverty puts a strain on a marriage, creating tensions and dissatisfactions that contribute to a subsequent dissolution.[5]

Against this apparent consensus are one empirical finding and three theoretical arguments. First, the time trend is not supportive of the idea that rises in income are associated with increased marital stability. Second, consider the plausible hypothesis that in many instances income has a positive effect on divorce. After all, more income can make desired divorces and separate living arrangements affordable. Third, income is partly determined by personal traits that are themselves related to marital stability, so the empirical positive relation between income and marital stability does not imply a causal relation. Fourth, income from welfare may not be "ordinary income," because it may carry a stigma that is destabilizing.[6] Although no direct evidence indicates such a stigma effect is operative and destabilizing, that issue will be discussed later. To summarize, a negative income tax carries a direct income effect to the recipient married couple that is commonly viewed as promoting marital stability. We may consider marital stability to be, on average, a "normal good," but the evidence and theories appear only weakly supportive.

A second major difference is that AFDC requires the presence of dependent children for the receipt of transfer payments, whereas some proposed income maintenance plans, including those adopted in the Seattle-Denver experiment, do not. Because the presence of dependent children in a marriage has a negative effect on marital dissolution, this is an important difference in the two regimes. Restricting our attention to families with children will solve two problems: it will provide the proper comparison with existing AFDC programs, and also provide information about the only type of negative income tax legislation that is likely to be considered by Congress.

A third potential difference between the existing AFDC regime and the proposed negative income tax plan involves the level of payments received by a recipient family. To simplify matters, let us assume that the payment depends on three parameters: (a) the income guarantee for a family of a given size with no other income, (b) the benefit-reduction rate, and (c) the differential in income guarantees for families of different size and demographic composition. Regarding (c), let us assume that the negative income tax guarantee amounts are structured so that approximate neutrality in the economic well-being of different-sized families is achieved, if the guarantee were the only income received by the family. Given such a structure, the higher the income guarantee and the lower the benefit-reduction rate, the more generous is the plan, because more families are eligible and because recipient families will receive larger transfer payments.[7] In the following three comparisons of predicted effects on marital stability, let us assume that the mother maintains custody of the children.

1. *If the negative income tax and AFDC plans are equally generous* to a mother who is without her husband, then a net decrease in marital breakups is predicted. The main reason is that the price subsidy to the unmarried state is reduced. The mother would gain the same amount as before (under AFDC) if she is divorced or separated, and she would receive more than before if she remained married. The same statement applies to the husband-father, assuming that the negative income tax plan, like AFDC, provides no income support to the separated husband. If the husband were to become the provider of a new family and had a sufficiently low income, then he could gain under a negative income tax regime relative to AFDC. Let us assume that this possibility is sufficiently remote that the potential gain is negligible.

The negative income tax experiments, including Seattle-Denver, did differ from AFDC by providing an income guarantee and potential transfer payments to the departing husband, even if he remained single. This could be a major difference in the incentives to marital dissolution. However, the break-even levels of income for single persons were so low—around $2000 a year in 1970 dollars—that I will assume, unless otherwise noted, that the actual benefits to the husband from this provision of the negative income tax are negligible and can be ignored.

Returning to the payments received by the divorced mother, the term "independence effect" was used by Groeneveld, Hannan, and Tuma in their studies of marital dissolution to refer to the woman's opportunity to use these payments to support herself and her children. I prefer to speak of the price subsidy to being divorced and to focus on the neutrality or nonneutrality of the subsidy: the payments received when the woman is separated from her husband relative to the payments received if she stays married. (As noted above, the price subsidy also affects the husband by reducing his subjective obligation to pay child support or alimony.)

A second reason for fewer marital dissolutions under the negative income tax regimen relative to AFDC is that the married couple may receive income transfers. Here the focus is on the receipt of income, per se. Although the income effect does not have an unambiguous sign, prevailing opinion suggests that it should be weakly pro-stability.

In summary, if the negative income tax is as generous as AFDC, its price and income effects should lead to a reduction in marital splits. However, only a negligible decrease would be predicted by those who: (i) minimize the income effect; (ii) emphasize the fact that there has been no change in the independence effect; and (iii) believe that there are actually few cases under our current welfare system in which a father leaves his wife and children to permit them to qualify for transfer payments. A negative income tax plan of equal generosity obviates taking this drastic step, but apparently there has never been any concrete evidence that this behavior occurs.[8] Of course, the lack of empirical evidence does not mean that the events have not occurred.

2. *If the negative income tax is less generous than AFDC, and the two plans coexist,* then again a net decrease in marital breakups should occur, but the decrease should be even smaller than in the case of an equally generous negative income tax. In brief, the negative income tax again reduces the price subsidy to being divorced relative to being married, although the ratio of payments received with and without a divorce is smaller. Also, the presumed pro-stability income effect is smaller than in the previous case, because the transfer payments to the married couple are lower. Again, the potential transfer payments to the separated husband under a negative income tax plan probably play no role in his decision.

3. *If the negative income tax is more generous than AFDC,* then the two theoretical effects we have considered have opposite signs with respect to marital stability. The higher payments of the negative income tax would dominate AFDC, and the latter would disappear for lack of customers. Under the more generous negative income tax the payment to the divorced woman is increased, but so is the payment to the woman (and her husband) if she remains married. It is likely that the higher level of payments to the woman if she divorces dominates the comparisons in her decision to remain married or become divorced. In this sense, the "independence effect" of the income maintenance system has been increased, implying that the effective price subsidy to being divorced is increased, leading in turn to an increase in marital dissolutions. A more generous plan increases the payments to married couples, and the pure income effect, which is the second theoretical effect, may be assumed to promote marital stability. Thus, a more generous negative income tax induces a net price change that promotes marital breakups and an income change that promotes marital stability. For reasons discussed above, the price effect appears, a priori, to be the stronger.

A Broader Research Agenda for Marital Stability

The economic framework presented in the previous section was overly simplified in several respects. A more realistic setting and objectives will be useful for the reanalysis of the Seattle-Denver study and will reinforce the message of the difficulty, perhaps intractability, of social experiments on family issues.

A fundamental question that has been only partially addressed by past research is the precise purpose of analyzing marital dissolutions and other demographic outcomes of the experiments. Given our attention to welfare reform, one important purpose of an experiment is to measure the fiscal costs of the reform. The costs will tend to rise or fall depending on whether the reform increases or reduces marital dissolutions. Unlike reductions in labor supply and earnings, however, the change in dissolutions may have no effect on national income. For example, a married couple that is not receiving welfare benefits may split, after which the mother and children begin to receive welfare payments. The divorced husband may continue working as much as he did before, and the divorced wife may work in the market no less, and perhaps even more, than she did before. The change in marital stability has no clear effect on national income, even though its relation to program costs is useful to measure.

We would also like to know how the well-being of low-income married couples is affected by a change in marital dissolutions. With the introduction of a negative income tax, married couples have an expanded set of options regarding their living arrangements, and if they choose to change them, we may presume that they are better off. This is, of course, an application of the economist's conventional assumption of consumer sovereignty and rational behavior. The important point here is that those who use the experimental results to design welfare programs are not going to be able to answer these questions about well-being.

Finally, what is the effect on the well-being of the children in the low-income families that experience a change in marital status as a result of the negative income tax? This difficult question may require many years to elapse before it can be answered. However, the social concern about deleterious effects of marital breakups on the well-being of children is sufficiently widespread that we may agree on the importance of using the experiment to measure marital dissolutions in families with children.

Given the two purposes of measuring program costs and marital dissolutions in families with children, we are in a firmer position to assess the strengths and weaknesses of different experimental designs. A well-known weakness of the experiments, in terms of measuring the demographic consequences of a nationally legislated negative income tax, is their short duration—three to five years. Short-duration experiments do not simulate the incentives of a permanent negative income tax pertaining to demographic behavior. The apparent bias in a short-duration experiment would be to understate the program's effect on the lifetime or steady-state incidence of births, marriages, and divorces, since the present value of the subsidies from a short-duration program are lower. However, the timing of these outcomes may be so affected that a short-duration experiment could overstate, rather than understate, the impact of a permanent program.

To illustrate these biases, let us use the example of births. With a permanent negative income tax, a married couple might decide to have three instead of two children in response to the incentive of the extra transfer payments they will receive during the 16 to 20 or so years that the child is their dependent. Another incentive is that any reduction in market earnings by the mother during the first 10 or so years of the additional child's life may also be partially offset by an additional increase in transfer payments. Clearly, the total value of these extra transfers under a permanent negative income tax is considerably larger than the payments received under a short-duration experiment. Thus, the lifetime incidence of the births of additional children may be substantially understated in the experiment. But to illustrate an opposite bias whereby fertility is overstated, consider all couples who plan to have an extra child and who would do so whether or not a negative income tax program exists. They might respond to the subsidy for births in the short-duration experiment by bearing that child "now" rather than "later." The short-duration experiment will, in these cases, overstate the fertility effect of a negative income tax.

It is partly a matter of judgment and partly a matter of ingenuity in analyzing the experimental data to determine which behavioral outcomes are affected by these duration biases and to measure the bias. Regarding marital dissolutions, we might suspect that teenage cohabiting couples who did not have a legal marriage and had no children would be more likely to alter the timing (as well as the incidence) of dissolutions in response to financial incentives than would legally married couples in their thirties with children present.

Another shortcoming of the negative income tax experiments regarding the measurement of marital stability is the reliance on already married couples. There are two problems here. One is that the relevant population, given our interest in children in families without a father present, includes women who have never married and some who are divorced or separated and are not now married. A second and related problem is that the measure of marital breakups with a sample of already married couples may be biased even as a measure of dissolutions among married couples. Both problems arise because a permanent negative income tax may be expected to affect the decision to marry and to remarry, as well as the decision to dissolve an existing marriage.

In considering the problem of the relevant populations in connection with the well-being of children, an example may illustrate some important issues. Let us define family stability in terms of the proportion of time that children spend growing up with their mother and father present. So defined, an increase in marital dissolutions could be consistent with an *increase* in family stability according to the following scenario. Assume that a negative income tax that covers married couples increases the proportion of young unwed mothers who marry the fathers of their children. Assume further that these unwed mothers would not have married in the absence of the program. These two assumed outcomes are realistic because the current AFDC program, which provides transfer payments to the mother only if she does not marry, is assumed to be superseded by a negative income tax that provides transfer payments to the mother if she does marry (and is income eligible). Finally, assume that the proportion of these marriages that ends in divorce is higher than the proportion of divorces in the rest of the married-couple population. (This is also a realistic assumption.) The end result is that the overall proportion of marital dissolutions is increased. Nevertheless, the assumed marriages of unwed mothers who would not have married otherwise increases family stability as defined by the presence of a father and mother during the time of the upbringing of their children. The same apparent paradox—an increase in family stability accompanied by an increase in marital dissolutions—will result from a similar scenario applied to remarriages of divorced mothers with children.

The above example illustrates the point that female headship is probably more influenced by the current welfare system than is marital stability. AFDC may not create female-headed families by providing a monetary incentive for a father to leave his family nearly as often as it does by discouraging the marriage of young unwed mothers and of currently divorced and separated women who are receiving AFDC and other welfare assistance.[9] Thus, the proportion of dissolved marriages among women who marry could be *reduced* by the current system because the system discourages certain marriages from occurring in the first place.

Now consider a second bias in measuring marital dissolutions that arises when examining only the existing stock of married couples.[10] Two examples will illustrate the problem. First, assume that unmarried women tend to have preferences in favor of singleness and against marriage. If a negative income tax offsets the current incentive to singleness and encourages more marriages among these women, we might expect that the proportion of divorces in these marriages would be above average. As a consequence, the long-run impact of a negative income tax on divorces is understated by observing only the already married population.

An opposite bias is also possible. Assume that the population of unmarried women is composed of two groups: one that is committed to singleness and will not marry and another that is planning on marrying but is taking more time to search more carefully. If a negative income tax encourages more, or even less delayed, marriages among this second group, we might expect fewer divorces among these marriages than the average. If so, then the full impact on divorces is overstated by using only the already married population in an experiment.

The data from the negative income tax experiments do not realistically allow analyses to estimate the effect on (a) first marriages of never-married women; (b) marriages of unwed mothers; (c) remarriages of divorced women who are not currently married; and (d) subsequent marital decisions. One problem is the small sample sizes of some of the populations. A second is the short duration of the experiment relative to the time horizons for these decisions. Again, the downward bias of a short-duration plan, stemming from the lower present value of the plan's transfer payments, is competing with the upward bias from intertemporal substitution.

The case of new marriages by women and men without children is worth special attention to reveal some of the complexities in using the experiments to analyze marital behavior. The Seattle-Denver experiment created unusual bonuses for new marriages. Imagine an 18- to 20-year old unmarried son or daughter in an experimental family who is considering marrying or cohabiting. The first bonus to marriage is eligibility for cash transfers for the couple, including an additional transfer allowance that the experimental plan will assign to the new partner. These cash transfers are not available to *either* partner under the existing welfare system. It is also unlikely that a couple without children would be eligible for cash transfers in a nationally legislated plan. The son or daughter in a family eligible to receive experimental negative income tax payments received a dowry to a new union unavailable to other people in that "marriage market," and operative only for the duration of the experiment. A second bonus to the new marriage or cohabitation was that eligibility for experimental payments was extended to the new partners even if they later dissolved the union and formed a second union with a different partner.[11]

My point is not to dwell on the difficulties in using existing experimental data to test for effects of a negative income tax on new marriages and subsequent dissolutions. Rather, it is that the already married couples are the only feasible group to use to examine marital stability and that there are inherent limitations in relying on this group.

Summary Points on the Experiments and the Issue of Marital Stability

Several lessons can be drawn about the use of experiments to study marital stability. First, the population of interest should include the major groups that will be affected, and in particular, young persons who are not married. Second, long-duration experiments seem necessary both because the decisions about marital status and other family issues involve long-duration plans and consequences and because the biases from using short-duration experiments are not clear in direction. Both lessons would apply to the issues of fertility and marriage as well as to marital dissolution. A third lesson, which is derived from our interest in welfare reform, is that families with children should receive priority in the design of the experiment, which should include a scheme in which families without children are not eligible for payments. Finally, a simplified economic framework for analyzing how a negative income tax influences marital dissolution suggests two predictions:

- A negative income tax plan that is as generous as an AFDC plan or less generous should promote marital stability.
- The predicted effect on marital stability of a plan that is more generous than the existing AFDC plan is ambiguous. If the negative income tax led to fewer marital splits, we could infer that the gain to a married couple that stays together dominates the extra gain that the mother would receive from the new plan if the marriage dissolves. If the negative income tax led to more marital splits, then that latter gain would appear to dominate the decision.

The Experiments, with Special Reference to the Seattle-Denver Experiment

The findings on marital stability that received by far the most attention are those from the Seattle-Denver experiment, which was the largest of the four negative income tax experiments. Seattle-Denver had the advantage of a five-year duration for a subsample of about 25 percent of the married couples; most of the other 75 percent were in the experiment for three years, as were the subjects in the other three experiments. In fact, a small subsample of 169 families was assigned to a third, 20-year duration, category. This assignment occurred after the experiment had started and after these families had been originally assigned to three-year treatment and control groups. This 20-year group was not analyzed separately in the research on marital issues, but will be referred to later.

The Seattle-Denver experiment was the last to be completed, and its research team of Groeneveld, Hannan, and Tuma reviewed and reanalyzed the findings on marital stability from the other three experiments. They concluded that these, individually or collectively, did not show any clear impact.[12] The sample size in the rural experiment was too small for conclusive evidence, particularly in view of the low proportion of families experiencing a marital dissolution—around 2 percent for both treatment and control groups. The Gary experiment showed no effect on marital splits, but the Groeneveld, Hannan, and Tuma team pointed to administrative and data flaws in this experiment that led them to discard these results. The first experiment, carried out in New Jersey and Pennsylvania, appeared to have inconclusive results regarding marital splits when originally analyzed, but the researchers reanalyzed the data and found evidence for a pro-split outcome that supported the conclusions of their Seattle-Denver research. The New Jersey study had a much smaller sample and more attrition among married-couple families than did Seattle-Denver, however.

In the Seattle-Denver experiment, marriage was defined by cohabitation, not by a marriage certificate or other legal sanction. A marital dissolution was defined by a separation, not necessarily a divorce, of at least 30 days and a statement from one of the partners that a separation had occurred.[13] Because this reanalysis is restricted to couples with children, the problem of dealing with unmarried cohabiting couples is presumed to be negligible. The Seattle-Denver definition of a separation seems to set rather loose criteria, however, and the effects of this on reports of a dissolution by experimental couples relative to control couples will be discussed below.

Several features of the Seattle-Denver experiment appear either to be obstacles to analyzing marital dissolutions or to imply reservations about the research findings.

1. The addition of a training and counseling program. A large proportion of the experimental group was given the option of a training and/or counseling program (hereinafter referred to as the training program) intended to improve the earning capacities of the adult family members. In fact, the Seattle-Denver experiment consisted of four major groups: families that were assigned only to a training treatment, families assigned only to a negative income tax plan, families assigned to both treatments, and control families. The training program complicates measuring the effect of a negative income tax on marital stability in three ways:

- No theoretical basis exists for predicting the sign of the effect of training on marital stability. The effect may differ depending on whether the husband or the wife receives the training.
- The sample size for the "pure" negative income tax treatment is sharply diminished.
- Training programs are difficult to administer in a way that will replicate how a nationally legislated program would be carried out and will be as unobtrusive to the experimental subjects as a nationally legislated plan. In contrast, a negative income tax plan has relatively rigid parameters that permit the experimental plan

to match closely the design of a nationally legislated plan, and its administrators play a relatively passive role.[14]

2. Small sample sizes for subgroups of interest. The Seattle-Denver experiment was directed toward three major ethnic groups: white, black, and Hispanic, the last primarily of Mexican heritage and living in Denver. There were eleven different negative income tax plans, two (or three) durations, and three training treatments. Problems of inadequate sample sizes arise when these features are extensively cross-classified.

3. The short duration of the experiment. The short duration of the experiments relative to the time horizons of such demographic behavior as marriage and divorce led Groeneveld, Hannan, and Tuma to emphasize the five-year plan, on grounds that it ''is more like a permanent program than the 3-year treatment.'' In reporting that the three-year treatment was about 75 percent as large as the five-year treatment, they commented that ''if the longer treatment more closely approximated the effects of a permanent program, a permanent program would have even larger effects than the 5-year program.''[15] These arguments imply that the intertemporal substitution bias is dominated by duration bias; that is, the five-year plan's lesser duration bias gives it an overall larger effect. However, the reanalysis of the Seattle-Denver data presented below does not support this finding.

4. Attrition. All research with longitudinal surveys has to deal with attrition bias. The attrition proportion was about 20 percent in the negative income tax experiments, including Seattle-Denver,[16] and the resulting biases may be serious. For example, the attrition bias of most concern in analyzing labor supply is that families whose earnings *declined* had an incentive *not to drop out* if they were in the experimental group, but had no such incentive if they were in the control group. Thus, the experimentals who did not drop out should overrepresent experimentals whose earnings declined, especially those who lost their jobs and had zero earnings. The controls had no such incentive not to drop out, and controls did drop out more than experimentals.

The attrition bias might be even more serious in analyzing marital stability than it is for labor supply analysis. A decline in earnings associated with decreased labor supply will consist of a continuum of small to large declines, whereas the decline in income for the mother after the departure of her husband is often a very large loss. A woman in the negative income tax experimental group whose marriage breaks up has the option of receiving the higher of the experimental payment or the AFDC payment. Women in the control group whose marriages dissolve get nothing from the negative income tax before or after the marital dissolution, so they have no economic incentive to continue responding to the interviews three times a year. Attrition rates were higher for control wives than experimental wives in the Seattle-Denver experiment, and the ratio of attrition rates, control-to-experimental, was higher for wives than for husbands. Experimental husbands were, of course, less likely to receive transfer payments after a marital dissolution than were their wives. The economic incentive was also evident within the experimental group: wives assigned to the most generous plans had lower rates of attrition than wives in the least generous plans.[17]

In summary, the sample of experimental families who remained in the study should overrepresent marital dissolutions relative to the full original sample of experimentals. The economic incentive for this attrition bias is not present in the control group, and other things equal, we might view the proportion of marital dissolutions in the remaining control group as an unbiased estimate of the proportion for the full sample of controls. However, if we believe that personal problems and traits that are associated with marital dissolution are also associated with attrition, then, with no economic incentive in operation, those controls who dropped out might well have a higher incidence of marital splits.[18]

5. Post-experimental design changes and the 20-year treatment. After the experiment began, the initial design was changed in several ways. First, shortly after the experiment began about 40 percent of the control families were assigned to a five-year control status.[19] The five-year controls were thus exempt from the substantial attrition that occurs early, and for this reason their attrition was much lower than that of the three-year controls. To the extent that the frequency rates for marital splits are affected by attrition bias, the rates for the five-year controls will differ from those for the three-year controls.

Another change in design occurred after the experiment had been in operation for about two and one-half years, when 169 families in Denver were assigned to a 20-year negative income tax plan: 112 families that were initially control families, and a second group, reassigned a few months later, of 57 families that were originally three-year experimentals. The latter maintained their originally assigned guarantee amount and benefit-reduction rate.[20]

Several complications arise from these reassignments, which are not discussed here.[21] As shown below, dealing with these complications is facilitated by the statistical techniques presented by Groeneveld, Hannan, and Tuma, which make use of multiple time periods for each couple. This allows the couple's particular and varying experimental status to be matched with the time period under investigation.

6. Fraud, reconciliations, and reporting biases. Problems associated with fraud, reconciliations, and reporting bias are all somewhat related. The issues raised can become complicated, but neither the Seattle-Denver researchers nor I view the observable evidence as indicating a serious bias from these sources. The clearest case of fraud is where an experimental couple falsely claims to be separated in order to collect

extra payments. Then the two main objectives of the experiment will suffer from a bias. Obviously, the true number of marital dissolutions will be overstated. The cost of the program is also overstated, assuming the amount of fraud in the experiment exceeds the amount that would exist in a nationally legislated plan. Reconciliations that occur during the experiment or soon after the experiment ends might be an indicator of this type of fraud. Reconciliations are also of concern because they affect the time in which children have both parents present.

Finally, the reporting for experimentals and controls differed in three ways. Controls reported their family composition once every four months, whereas the experimentals reported monthly as well as in triennial interviews. Groeneveld, Hannan, and Tuma dealt with this problem by using reported dissolutions from the same triennial interview. However, this does not equalize the incentives the experimental families have to report a dissolution or the likelihood that the multiple sources of reports by these families would spill over to their interviews.[22]

In summary, the features of the Seattle-Denver experiment that I find most likely to lead to important biases about marital dissolution in relation to a negative income tax are (a) the confounding of negative income tax treatments with training treatments; (b) the duration biases; and (c) attrition. The problem of sample size is one of reliability, not bias. The problem of the 20-year treatments is mainly that they create ambiguity in interpretation. Fraud, reconciliation, and reporting issues appear minor, but it is worth noting that the *direction* of bias is surely that of exaggerating the effect of a negative income tax on marital splits.

The Seattle-Denver Results Concerning Marital Dissolution

In their final report on marital dissolutions, Groeneveld, Hannan, and Tuma state that "the NIT plans tested in SIME/DIME dramatically increased the rates at which marriages dissolved among white and black couples . . ." They report an increase of "40 to 60 percent," and add:

> If one wishes a single set of numbers to summarize our findings one might choose the effects of the $3800 guarantee level treatments because it is closest to the current system and to likely welfare reforms. [Adjusting for attrition bias and restricting the estimates to couples with children] one obtains estimates of 58 percent increases in dissolution for blacks and 51 percent increases for whites. These are estimates of the experimental-control differences in the SIME/DIME population for the most feasible programs tested.[23]

As large as these increases are, they are smaller than those reported by these researchers in earlier published articles. In the first article the $3800 guarantee plan was estimated to increase "the annual probability of marital dissolution . . . by 63 percent for blacks, 194 percent for whites, and 83 percent for chicanas over what it would be in the control situation.''[24] These results applied to the first two years of the experiment. For the first three years of the experiment the estimates of increases ranged from 57 percent to 129 percent (in a 1979 article) and from 24 percent to 114 percent (in a 1980 article).[25] In the latter article the results for Chicanos showed no increase in marital dissolutions for the treatment group.

The increase in marital splits among experimentals relative to controls was not attributable to a low proportion of splits in the control group. The proportion of white, black, and Chicano couples in the control group who experienced a marital dissolution during the first three years were 16, 24, and 20 percent, respectively. These percentages apply to the originally enrolled couples who did not drop out and they reflect the full three years of exposure to risk. The percentages are considerably higher than those reported by Sawhill et al. for poor couples in the Survey of Income Dynamics for a similar time period[26] or for comparable controls in the New Jersey negative income tax experiment.[27] Because the dissolution proportions among the controls in Seattle-Denver were high, the even higher level of dissolutions among the treatment group was noteworthy.

The Seattle-Denver results were surprising in two respects. First, previous research on the impact of AFDC on marital dissolutions had not prepared researchers to see a large effect from a negative income tax. After all, no firm evidence had been established for a large destabilizing effect of AFDC on marriages despite the fact that the system essentially provided ''permanent'' benefits to a wife if her marriage dissolved and no benefits to a married couple.[28] The Seattle-Denver experiment showed a large destabilizing impact from a program that did provide benefits to a couple that stayed together.

One possibility could have been that the destabilizing effect was attributable to the relative generosity of the negative income tax plans. In other words, the price effect was so large in its negative impact on stability that it dominated any positive income effect. Actually, this was not the case. In the second surprising result, the least generous negative income tax plans, which offered about the same or lower cash payments as did AFDC, induced the largest destabilizing effect, while the most generous plan had essentially no destabilizing effect. This is the opposite of the theoretically expected result discussed earlier.

The researchers rationalized the large destabilizing impact of the low-payment plan by suggesting a negligible income effect associated with the payments to the intact couple, while emphasizing a large price or ''independence'' effect of the payment to the divorced mother.[29] (The relative sizes of the income and independence effects are claimed to be reversed for the high-payment plan.) Although the ostensible payment to the divorced mother from the low-payment plan is no more than that available under AFDC, Groeneveld, Hannan, and Tuma suggest that the negative income tax payment is worth more because it carries less stigma, it is more certain to be received, and it involves low transaction costs because the woman does not have to file and wait for AFDC benefits.[30]

These reasons do not appear persuasive as explanations for a steady-state increase in ''permanent'' dissolutions, defined here as those dissolutions that prevail for many months. Instead, the explanations appear to apply to the timing of the dissolution rather than to its eventual incidence. The low transactions costs, for example, should only affect the timing. The fact that the woman may receive immediate monthly payments from the negative income tax plan increases the present value of the total payments received. However, this increase is trivial relative to the present value of AFDC payments because the latter are ''permanent'' and include noncash benefits, whereas the negative income tax cash payments will terminate within a year or two for most three-year plans and within three years or so for most five-year plans. Similarly, the certainty of the negative income tax payments should be important, if at all, only with respect to the timing of the marital dissolution.[31]

We do not have direct evidence for a stigma effect that discounts AFDC payments, and if we did we would need to know how a legislated negative income tax plan would enroll and monitor its participants to determine whether its administration would eliminate any stigma in receiving payments.[32] Apparently, many of the experimental families who were already receiving AFDC in Seattle-Denver were unwilling to shift to the negative income tax plans even when the latter paid larger cash transfer payments. These AFDC recipients did not want to jeopardize their Medicaid benefits or, in some cases, housing subsidies.[33] Persons already on AFDC may be inured to stigma, but their reluctance to shift to higher-paying negative income tax plans casts doubt on the strength of the stigma effect. Again, the stigma of AFDC might delay a woman's shift from receiving negative income tax payments to receiving AFDC benefits when the latter are larger, but this behavior implies that the negative income tax plan is affecting only the timing of the split, not its incidence.

Expressing these doubts about the explanation Groeneveld, Hannan and Tuma offer for their surprising results does not refute their explanation. Indeed, rather than attempt a thorough analysis of their explanations, the next section presents a reanalysis of their data.

An Empirical Reanalysis of the Seattle-Denver Data

The empirical reanalysis of the Seattle-Denver data presented in this paper will concentrate on couples with children and on marital dissolutions as the outcome of interest.[34]

The techniques of analysis follow closely the pioneering use by Groeneveld, Hannan, and Tuma of "event-history analysis," which appears preferable to any other statistical procedure for summarizing the results. These techniques focus attention on the *rate* of dissolution; that is, the sample's proportion of dissolutions per unit of time. The time-unit may be a year or as brief as a day, because the calendar date of the dissolution is recorded in the Seattle-Denver data. Remarriages, reconciliations, and subsequent dissolutions are not analyzed. The first dissolution ends the couple's record.

An important reason for using the rate instead of just measuring the incidence of a dissolution is that the treatment and control groups are exposed to the risk of dissolution for varying lengths of time. Even groups in the experiment for the same intended duration, whether three, five, or 20 years, may experience differential attrition. In particular, more attrition on the part of the control group could yield a spuriously lower incidence of dissolutions, and this bias would be all the greater if the control couples were more likely to divorce or separate after they dropped out of the experiment.

What is less obvious, however, is that the rate measure may bias (or exaggerate) marital dissolutions of treatment couples relative to control couples in the context of a short-duration experiment. As discussed earlier, the short-duration experiment provides an artificial incentive to divorce earlier rather than later. Previously, this intertemporal substitution bias was cited as a reason why the incidence of dissolution during a three-year experimental period might be higher than the incidence for the same three-year period under a permanent plan. The rate measure could increase this bias because even the same number (incidence) of dissolutions in a three-year period will produce different rates—a higher rate when the dissolutions occur early.

Table 1 illustrates these distinctions between rates and incidence, early and late dissolutions, and records with and without attrition. A hypothetical example of four couples (A,B,C, and D) and three periods is shown. Case II relative to Case I shows that later dissolutions yield a lower rate for the same incidence level. Case III relative to Case II shows how attrition will tend to understate the dissolutions if an incidence measure is used, whereas the rate will adjust for the varying exposures to risk.

Case IV relative to Case I is interesting because it reveals how the same rate may accompany different levels of incidence. Note that although Case I has the same rate as Case IV, one dissolution per four time-periods of exposure, there are only two dissolutions among the four couples in Case I and three dissolutions in Case IV. Our concern about the disruption of intact marriages and the consequences of this for the upbringing of children probably implies that Case IV is "worse." The important point is that a short-duration experiment should tend to produce Case I-type outcomes among the treatment group.

Table 1
Illustrative Examples of Differences in the Incidence and Rate of Marital Dissolutions

Legend: 1 = marital dissolution
0 = no marital dissolution
X = post-event period (no further record)
ATT = attrited (no information)

(Data beyond three years are unavailable.)

Period	Couple				
	A	B	C	D	
Case I					
1	1	1	0	0	Incidence: 2/4 = .50
2	X	X	0	0	
3	X	X	0	0	Rate: 2/8 = .25
Case II					
1	0	0	0	0	Incidence: 2/4 = .50
2	0	0	0	0	
3	1	1	0	0	Rate: 2/12 = .17
Case III					
1	0	0	0	0	Incidence: 2/4 = .50
2	0	0	0	ATT	
3	1	1	0	ATT	Rate: 2/10 = .20
Case IV					
1	0	0	0	0	Incidence: 3/4 = .75
2	0	0	0	0	
3	1	1	0	1	Rate: 3/12 = .25

A First Look, Using The Seattle Data

The yearly records of the Seattle experiment may be shown with three tables similar in form to Table 1. This will permit us to see the ingredients of analysis that will later be summarized in a statistical model, and the simplicity of the tables will facilitate some important observations about the data. Table 2 shows the number of couples and their record of attrition from the experiment. Tables 3 and 4 show the year-by-year record of marital dissolutions for white and black couples respectively.

Assignments to the four experimental groups were random within the stratifications of city, ethnicity, and the estimated level of the families' normal incomes.[35] As we shall see, income is not a major determinant of marital dissolution in this sample, so ignoring this variable in these tables still allows a fairly accurate picture of marital dissolutions in Seattle in response to the experimental plans. Adding the income variable would further dilute the already thin cell sizes. The five-year experimental cells are particularly small.

Seattle data are easy to interpret because the sample was stratified with only two ethnic groups and two duration groups. No family units were shifted among plans as they were in Denver. However, Seattle's outcomes are quite different from those in Denver in certain key respects. An overall assessment must wait for the statistical model for Seattle and Denver combined.

Table 2 shows Seattle's number (N) of originally married couples and the number who dropped out of the experiment, for each race for the four experimental statuses and the two duration groups. Overall, the attrition rate is 16 percent for the entire number (163 couples out of 1001), and the rates for the groups are: 20 percent for controls, 17 percent for the trainee group (TR), 15 percent for the pure negative income tax (NIT) group, and 13 percent for the group receiving the combined treatment of a negative income tax and training (NIT × TR). A striking

Table 2
Attrition in Seattle Negative Income Tax Experiment, by Treatment, Race, and Duration of Assignment

Racial Group and Duration of Experiment	Control			Training			NIT			NIT x Training		
	N	ATT	R	N	ATT	R	N	ATT	R	N	ATT	R
White												
3-year	103	19	.18	96	12	.12	79	9	.11	148	18	.12
5-year	69	4	.06	33	4	.12	35	3	.09	45	5	.11
Total	172	23	.13	129	16	.12	114	12	.11	193	23	.12
Black												
3-year	47	24	.51	57	11	.19	50	8	.16	109	14	.13
5-year	44	6	.14	28	9	.32	25	9	.36	33	8	.24
Total	91	30	.33	85	20	.24	75	17	.23	142	22	.15
Total												
3-year	150	43	.29	153	23	.15	129	17	.13	257	32	.12
5-year	113	10	.09	61	13	.21	60	12	.20	78	13	.17
Total	263	53	.20	214	36	.17	189	29	.15	335	45	.13

Notes: N = number of couples at beginning of experiment. ATT = Attrition (number of couples who dropped out). R = rate of attrition (percentage dropping out). Table refers to originally married couples with children under age 21 at beginning of experiment. Cases where a spouse died during the experiment have been excluded.

finding is that the controls who were in the three-year duration group had a rate of attrition about three times that of the five-year controls (29 percent compared to 9 percent, for whites and blacks combined), despite the fact that the five-year controls had two additional years of exposure to the risk of dropping out of the experiment. The reason is surely that the five-year controls were assigned after the experiment had begun and after the considerable attrition in the beginning months of the experiment had occurred. Clearly, the designation to three-year and five-year control status was nonrandom with respect to attrition. In all subsequent analysis, all the controls in each year of their participation in the experiment are pooled to guard against the assignment being nonrandom with respect to the propensity to divorce or separate. Together they should constitute a random group.

Tables 3 and 4 show the essential information on marital dissolutions for white and black couples. For each year and for each experimen-

Table 3
Annual Rates of Marital Dissolution among Whites in Seattle Experiment[a]

Duration and Year in Experiment	Experimental Treatment: Control[b]			Training			NIT			NIT x Training		
	N	D	R	N	D	R	N	D	R	N	D	R
3-yr, first	166	10	.060	93	7	.075	78	5	.064	143.5	10	.070
3-yr, second	149.5	9	.060	80.5	6	.075	71	7	.099	128	16	.125
3-yr, third	136	12	.088	72.5	4	.055	60.5	5	.083	107.5	6	.056
5-yr, first	included above			31.5	3	.095	36	2	.056	44	6	.136
5-yr, second	included above			27	1	.037	33.5	3	.090	37.5	3	.080
5-yr, third	included above			25.5	2	.078	29.5	1	.034	33.5	2	.053
5-yr, fourth	50	1	.020	24	1	.042	27.5	3	.109	31	0	.000
5-yr, fifth	49	1	.019	23	0	.000	24	0	.000	30.5	1	.033
Totals												
3-yr	451.5	31	.069	246	17	.069	209.5	17	.081	379	32	.034
5-yr	550.5	33	.060	131	7	.053	145.5	9	.062	176.5	12	.068
Total	550.5	33	.060	377	24	.064	355	26	.073	555.5	44	.079

	Control Dissolution Rate (Adjusted for Attrition[c])	Ratio[d] No Adj.	Ratio[d] Att. Adj.	Ratio No Adj.	Ratio Att. Adj.	Ratio No Adj.	Ratio Att. Adj.
3-yr	.070	1.00	.99	1.17[d]	1.13[d]	1.22	1.16
5-yr	.062	.88	.88	1.03	.90	1.13	.94
Total	.062	1.07	1.03	1.22	1.11	1.32	1.16

Notes follow Table 4. See also Notes to Tables 5 and 6
N = number of person- (couple-) years of exposure to risk. In the first year, N = the number of couples at the beginning of the experiment minus one-half of the number of couples dropping out who did not divorce or separate.
D = number of marital dissolutions.
R = rate of marital dissolution, measured as the proportion of dissolutions per years at risk.

Table 4
Annual Rates of Marital Dissolution among Blacks in Seattle Experiment[a]

Duration and Year in Experiment	Control[b] N	Control D	Control R	Training N	Training D	Training R	NIT N	NIT D	NIT R	NIT x Training N	NIT x Training D	NIT x Training R
3-yr, first	83.5	8	.096	54.5	4	.073	48.5	3	.062	105	23	.219
3-yr, second	65	6	.092	46.5	1	.022	43	3	.070	78.5	11	.140
3-yr, third	53	3	.057	42.5	2	.047	38.5	3	.078	65.5	4	.061
5-yr, first	included above			26	0	.000	23.5	2	.085	31.5	5	.159
5-yr, second	included above			23	4	.174	19.5	2	.103	24	3	.125
5-yr, third	included above			18.5	0	.000	16.5	0	.000	20.5	3	.140
5-yr, fourth	30.5	2	.066	17	1	.059	15.5	0	.000	17	2	.118
5-yr, fifth	28	1	.036	15	1	.125	13.5	1	.074	14	1	.071
Totals												
3-yr	201.5	17	.084	143.5	7	.049	130	9	.069	249	38	.153
5-yr	260	20	.077	99.5	6	.060	88.5	5	.056	107	14	.131
Total	260	20	.077	243	13	.053	218.5	14	.064	356	52	.146

	Control Dissolution Rate (Adjusted for Attrition[c])	Training Ratio[d] No Adj.	Training Ratio[d] Att. Adj.	NIT Ratio No Adj.	NIT Ratio Att. Adj.	NIT x Training Ratio No Adj.	NIT x Training Ratio Att. Adj.
3-yr	.088	.58	.56	.92	.86	1.82	1.67
5-yr	.082	.78	.73	.73	.63	1.70	1.48
Total	.082	.69	.65	.90	.74	1.90	1.70

[a]See also Notes to Tables 5 and 6.
N = number of person- (couple-) years of exposure to risk. In the first year, N = the number of couples at the beginning of the experiment minus one-half of the number of couples dropping out who did not divorce or separate.
D = number of marital dissolutions.
R = rate of marital dissolution, measured as the proportion of dissolutions per years at risk.

Notes to Tables 3 and 4

[a]Tables refer to originally married couples, with children under 21 at the beginning of the experiment; cases where spouse died during the duration of the experiment are excluded. Dropouts contribute to the total dissolution rate (or proportion) during the year that they leave the experiment by assuming they represent one-half year of exposure to risk. If they report a dissolution, the dissolution is included in the total number of dissolutions for that year.

[b]All controls are aggregated during the first three years of the experiment. Only the five-year controls are measured during the fourth and fifth years of the experiment.

[c]The attrition adjustment has two parts. First, the dissolution rate is assumed to be 25 percent greater among the control dropouts for whom no information is available. Second, the dissolution rate is assumed to be 50 percent less for experimental (treatment) dropouts who were assigned to a negative income tax plan and for whom no information is available. No attrition adjustment is made for the training-treatment group.

[d]From Table 3, the calculation of the two ratios for the NIT three-year sample of whites is demonstrated as follows:

1.17 = .081/.069 = NIT average rate for the three-year group / control average rate for the three years.

1.13 = .079/.070 = The attrition-adjusted dissolution rate for the NIT group, assuming the NIT dropouts with no information about dissolutions have a dissolution rate one-half as large as the remaining NIT couples / adjusted rate of control attrition.

All other ratios in Tables 3 and 4 are derived in the same way.

tal group, the number of dissolutions is recorded along with the number of person-years (or couple-years) at risk. A couple that drops out of the experiment in any year is assumed to provide a half-year of exposure to the risk of dissolution. A dissolution recorded for a couple that drops out in a year is counted in that year. In the following year, only the continuing and still-intact couples are at risk. The three-year and five-year treatment groups—Training (TR), Negative Income Tax (NIT), and NIT × TR—are separately recorded. The controls are pooled in the first three years, but only the five-year group is recorded for the fourth and fifth years of the experiment.

Several interesting results in Tables 3 and 4 will be shown to hold up in the final analysis when all data are used and when a number of exogenous control variables are held constant statistically.

1. Looking at the average yearly dissolution rate and the ratio of these averages to the corresponding control group, no consistent pattern emerged regarding the three-year groups versus the five-year groups or regarding three of the four experimental groups—control, Training, or the NIT group.

2. The experimental group, NIT × TR, shows a higher dissolution rate, and among blacks, the higher rate for the three-year experimentals is statistically significant at conventional levels.[36]

3. The cell sizes are too small to detect a time trend in the dissolution rates, although there is a hint of a downward trend in the experimental groups, as, for example, when the third year is compared with the first and second in the three-year groups, and the fourth and fifth years are compared with the first two years in the five-year groups. Intertemporal substitution will be examined below in more detail, although our full analysis of the time-dependence of dissolutions is not completed.

4. In the light of earlier findings by Groeneveld, Hannan, and Tuma, an unexpected result from the Seattle data is that the average annual dissolution rates among the three-year NIT and NIT × TR groups are *higher* than among their five-year counterparts. Moreover, the NIT/Control ratios of the three-year dissolution rates are higher than these ratios for the five-year dissolution rates. Again, intertemporal substitution is a possible explanation.

5. An adjustment for attrition bias can be demonstrated with these data and it turns out to be a fairly minor adjustment. Couples who dropped out and who did not report a marital split represent a certain number of subsequent unknown person-years. In sensitivity tests, the marital dissolution rate is assumed to be 25 percent higher among controls who dropped out; say, .075 per year instead of .06. When this adjustment is used, the *overall* average dissolution rate of controls is raised by .001 or .002, from, say, .06 to .061 or .062. In the next step, the

dissolution rate among dropout experimentals who were eligible for negative income tax payments is assumed to be 50 percent smaller than the rate among experimentals who stayed; say, .03 instead of .06. Applying this rate to the unknown person-years among experimentals who dropped out serves in practice to lower the overall average dissolution rate of the NIT or NIT × TR groups by .002 to .005. Thus, the attrition adjustment could change the experimental/control ratio of dissolution rates by around 5 to 10 percentage points; for example, from .06/.06 = 1 to .058/.061 = .95 or to .056/.062 = .90. These calculations merely illustrate the sensitivity of the estimates to an attrition adjustment. They will now be set aside until the concluding section of the paper.

The Full Sample and The Use of an Exponential Model

Table 5 provides a relatively complete summary of the dissolution "effects" of the various experimental statuses, using the full information for both cities, all five years of the regular experiment, and the sixth and seventh years for the small number of Denver couples assigned to the 20-year duration plans. Also, the full set of control variables used by Groeneveld, Hannan, and Tuma is included in the statistical model. The reported coefficients under the column headed b show the approximate percentage effects of the independent variables on the marital dissolution rate. The coefficients of the experimental categorical variables are related to the "multipliers" of the dissolution rate of the omitted base group of controls.[37]

In Table 5 the original numbers of couples for each group in each plan are shown in brackets, and we see the small number of families in the 20-year plans. All of these 20-year couples were originally in another group, so the total number of couples at the beginning of the experiment is given by the totals for the 3- and 5-year groups, along with the controls: 272 white controls, 182 black controls, and 93 Chicano controls. Hence, the number of observations per group may well be too small for the analysis of a relatively uncommon event like marital dissolutions. (Table 6 shows the statistical results when using fewer groups but with larger cell sizes.) The general lack of statistical significance also discourages spending much effort in investigating the effects of the still smaller subgroups of experimental treatments, such as the three training programs and the eleven (or even three) negative income tax plans. Groeneveld, Hannan, and Tuma extensively analyzed the results for high, medium, and low negative income tax plans.

The statistical model underlying the results shown in Tables 5 and 6 is the discrete-time analogue of the continuous-time model used by Groeneveld, Hannan, and Tuma; that is, their exponential rate model.

Table 5
Estimated Effects of Independent Variables on Dissolution Rates: Full Set of Interactions, Treatment x Duration x Site [a]

Independent Variable[c]	Whites[b] b	Whites[b] t-ratio	Whites[b] original n	Blacks[b] b	Blacks[b] t-ratio	Blacks[b] original n	Chicanos[b] b	Chicanos[b] t-ratio	Chicanos[b] original n
Constant	−2.02	(3.50)***		−1.37	(2.69)***		−1.31	(1.22)	
Normal earnings ($000's)									
0-1	−.03	(.08)		−.31	(.73)		.19	(.25)	
1-3	−.19	(.48)		−.29	(.82)		.54	(.73)	
3-5	−.42	(1.04)		−.31	(.86)		.48	(.64)	
5-7	−.67	(1.62)		−.46	(1.28)		.43	(.56)	
7-9	−.78	(1.77)*		−.10	(.27)		.64	(.81)	
9-13	−.63	(.78)		1.23	(2.13)**		−3.80	(1.12)	
Unreported	−2.83	(.85)		.37	(.35)		−3.74	(.59)	
Duration of marriage	−.08	(4.78)***		−.06	(4.09)***		−.04	(1.57)	
Wife's age	−.01	(.50)		−.03	(2.65)***		−.05	(2.12)**	
Wife's ed, 12	−.23	(1.71)*		−23	(1.68)*		−.40	(1.78)*	
Wife's ed, 12	−1.16	(2.08)**		.66	(1.41)		...	...	
Young Children	−.26	(1.40)		−.23	(1.39)		−.70	(2.36)***	
AFDC, pre	.30	(1.74)*		.04	(.20)		.70	(3.01)***	
TR x 3 x S	.21	(.73)	96	−.26	(.63)	57	...	...	...
TR x 5 x S	−.08	(.21)	33	−.11	(.25)	28	...	...	...
TR x 3 x D	.12	(.39)	83	.57	(2.19)**	80	−.12	(.36)	61
TR x 5 x D	−.41	(.81)	33	.08	(.22)	33	.20	(.50)	30
NIT x 3 x S	.19	(.69)	83	−.13	(.36)	50	...	...	...
NIT x 5 x S	.26	(.73)	35	−.29	(.61)	25	...	...	...
NIT x 3 x D	−.12	(.33)	57	.36	(1.00)	39	−.49	(1.32)	70
NIT x 5 x D	.06	(.20)	44	.88	(3.42)***	44	−.01	(.02)	30
NIT x 20 x D	.09	(.25)	35	.07	(.17)	23	−.53	(.96)	15
(NIT.TR) x 3 x S	.31	(1.36)	148	.73	(3.20)***	109	...	...	...
(NIT.TR) x 5 x S	.13	(.40)	45	.76	(2.43)**	33	...	...	...
(NIT.TR) x 3 x D	.20	(.85)	142	.74	(3.17)***	96	.07	(.26)	152
(NIT.TR) x 5 x D	−.02	(.05)	49	−.12	(.35)	39	−.26	(.83)	59

[a]See the text for a specification of the statistical model to estimate the rate of marital dissolution.

[b]b = Multiplier, approximately equal to the percentage effect of the independent variable on dissolution. See text footnote 37.

n = number of couples at beginning of experiment; DF = degrees of freedom, based on number of 6-month time periods per couple at risk (minus the number of independent variables). Whites: n = 1120, DF = 7120; Blacks: n = 815, DF = 4732; Chicanos: n = 495, DF = 2960.

[c]Independent variables are defined for their values at the beginning of the experiment:

- Duration of marriage in years.
- Wife's age, in years.
- Wife's ed (education): the category "less than 12 years of schooling" is the omitted category.
- Young children: 1 if a child under 6 years of age is present; 0 otherwise.
- AFDC, pre: 1 if wife had participated in AFDC in the year prior to enrollment; 0 otherwise.

TR x 3 x S = Training treatment only and 3-year duration and in Seattle. Other treatment statuses are defined accordingly.

(NIT.TR) = The combined treatment of an NIT plan *and* training.

*Statistically significant at the 10 percent level, two-tailed test.
**5 percent level.
***1 percent level.

Table 6
Estimated Effects of Independent Variables on Dissolution Rates: Summary Results, Combining Duration and Site Groups[a]

	Whites[b]		Blacks[c]		Chicanos[d]	
Variable	b	t	b	t	b	t
Constant	−1.92	(3.36)***	−1.30	(2.53)**	−1.30	(2.79)***
Normal earnings ($000)						
0−1	−.01	(.02)	−.37	(.88)	.12	(.16)
1−3	−.19	(.49)	−.34	(.96)	.49	(.66)
3−5	−.41	(1.02)	−.34	(.96)	.39	(.52)
5−7	−.66	(1.61)	−.48	(1.33)	.33	(.42)
7−9	−.78	(1.77)*	−.20	(.53)	.53	(.67)
9−13	−.58	(.71)	1.13	(1.98)**	−3.85	(1.14)
Unreported	−2.71	(.81)	.20	(.19)	−3.71	(.59)
Denver	−.16	(1.20)	.12	(.90)	...	...
Duration of marriage	−.08	(4.79)***	−.07	(4.29)***	−.04	(1.51)
Wife's age	−.01	(.59)	−.03	(2.79)***	−.05	(2.10)**
Wife's ed, 12	−.23	(1.72)*	−.20	(1.46)	−.41	(1.85)*
Wife's ed, > 12	−1.17	(2.10)**	.77	(1.66)	...	...
Young Children	−.25	(1.39)	−.26	(1.54)	−.69	(2.33)**
AFDC, pre	.29	(1.66)*	.00	(.03)	.66	(2.89)***
TR	.03	(.14)	.15	(.69)	.10	(.33)
NIT	.16	(.82)	.27	(1.28)	−.30	(1.00)
(NIT x TR)	.17	(.97)	.57	(3.06)***	−.01	(.03)

[a]See notes to Table 5. Denver = 1 if family lives in Denver, O if in Seattle.
[b]DF = 7129. (DF = Degrees of Freedom)
[c]DF = 4741.
[d]DF = 2964.

Define $P(t)$ as the probability that a couple experiences a dissolution at time t, conditional upon the couple being at risk at time t. The usual logit transformation of $P(t)$, related to a linear specification of explanatory variables, is:

$$\ln[P(t)/(1 - P(t))] = a + bx.$$

As the interval of time becomes smaller, the data approach continuous time. The specification of the dependent variable that provides an exact analogue to the continuous-time model is:[38]

$$\ln[-\ln(1 - P(t))].$$

Let $y = \ln[-\ln(1 - P(t))]$; T is a vector of treatment variables, and X is a vector of exogenous determinants of marital dissolution. The statistical model in Tables 5 and 6 has this double-log functional form and uses discrete data for six-month time periods:

$$y = T'a + X'\beta.$$

Estimation is by maximum likelihood logit analysis, using the GLIM statistical package.

Table 7
Estimated Effects of Independent Variables on Marital Dissolution Rates (using the same samples and variables as Groeneveld, Hannan, and Tuma)

	Black				White				Chicano			
	Cain		GHT		Cain		GHT		Cain		GHT	
	b	t	b	t	b	t	b	t	b	t	b	t
Constant	−.55	(.73)	.07	(.10)	−2.35	(3.44)***	−1.59	(2.30)	−1.76	(1.63)	−1.01	(.95)
Normal Earnings ($000's)												
0-1	.35	(.98)	.40	(1.11)	.78	(1.66)*	1.01	(2.30)**	.00	. . .	−.02	(.02)
1-3	−.13	(.36)	−.11	(.31)	.81	(2.68)***	.89	(2.87)***	−.19	(.43)	−.19	(.43)
3-5	−.17	(.76)	−.10	(.45)	.60	(2.31)**	.66	(2.44)**	.17	(.46)	.14	(.38)
5-7	−.31	(1.56)	−.28	(1.40)	.45	(1.86)*	.52	(2.00)**	.03	(.10)	.06	(.17)
7-9	−.37	(1.82)*	−.34	(1.70)*	.19	(.75)	.23	(.85)	−.05	(.14)	−.01	(.03)
9-13	−.53	(.52)	−.59	(.58)	1.28	(1.26)	1.37	(1.33)	. . .	. . .	7.52	(.11)
Denver	.28	(2.01)**	.28	(2.00)**	−.18	(1.28)	−.20	(1.43)	. . .	. . .	. . .	. . .
Dur. Marriage	−.05	(3.33)***	−.05	(5.00)***	−.09	(5.05)***	−.10	(3.33)***	−.04	(1.41)	−.03	(1.00)
Age-W	−.01	(.72)	−.01	(.50)	.01	(.42)	.01	(.50)	−.05	(1.61)	−.06	(2.00)
Ed-W	.00	(.08)	.01	(.20)	−.06	(1.89)*	−.08	(2.00)**	−.03	(.55)	−.03	(.60)
Age-H	−.03	(2.14)**	−.03	(3.00)***	−.02	(1.15)	−.02	(1.00)	.01	(.20)	.00	. . .
Ed-H	−.09	(2.61)**	−.08	(2.67)***	.01	(.31)	.02	(.67)	−.02	(.38)	.03	(.75)
Children, n	.07	(1.42)	.08	(1.60)	.04	(.75)	.05	(.83)	.12	(1.35)	.13	(1.44)
Young Children	−.24	(1.47)	−.29	(1.81)*	−.27	(1.67)*	−.29	(1.81)*	−.38	(1.35)	−.43	(1.54)
AFDC,	.04	(.22)	.05	(.26)	.45	(2.40)**	.50	(2.63)**	.61	(2.50)**	.67	(2.79)
M-1	.42	(2.00)**	.45	(2.25)**	.29	(1.40)	.32	(1.52)	.52	(1.88)*	.52	(1.93)
M-2	.24	(1.22)	.30	(1.50)	.15	(.74)	.14	(.70)	.12	(.41)	.13	(.45)
M-3	.25	(1.16)	.26	(1.24)	.34	(1.69)*	.33	(1.65)*	.22	(.67)	.18	(.56)
M, 5 yr	−.24	(.96)	−.38	(1.46)	−.15	(.57)	−.29	(1.07)	.00	. . .	−.04	(.10)
NIT	.41	(2.05)**	.45	(2.14)**	.36	(1.70)*	.43	(1.95)*	.05	(.17)	.01	(.03)
NIT, 3 yr	−.24	(1.05)	−.30	(1.30)	−.24	(1.02)	−.33	(1.38)	−.11	(.30)	.00	. . .

Denver = 1 if family lives in Denver; O if in Seattle.
Dur. Marr. = years married at beginning of experiment.
Age-W = age of wife; Age H = age of husband.
Ed-W = Wife's education (years) Ed-H = husband's.
Children, n = number of children.
Young Children = 1 if a child under six years of age is present; O otherwise.

M-1: least generous training program.
M-2: more generous training program.
M-3: most generous training program.
M-5 yr: if training subsidy variable is for 5 years.
NIT = Pure NIT and NITxTR pooled.
NIT, 3 yr. = 1 if family was in NIT or NITXTR experimental status and in the 3-year duration group; O otherwise.

NOTE: This table is a replication of Table 5.1.A in Groeneveld, Hannan, and Tuma, "Marital Stability," *Final Report*, p. 367. The GHT columns refer to a continuous-time model; the other columns refer to a discrete-time model.

Using time intervals of six months (instead of one year as in Tables 3 and 4), it is possible to replicate closely the results of Groeneveld, Hannan, and Tuma when using the same data. See Table 7 for the replication of their results for all originally married couples, including those without children, for the first three years of the experiment.[39]

The outcomes of the experimental plans shown in Table 5 are not easy to summarize. No treatment variables are statistically significant among white and Chicano samples, and imposing zero coefficients on all five variables defining any of the three experimental plans, TR, NIT, or NIT × TR, does not significantly worsen the fitted relation. In terms of the pure NIT plans, six of the 13 coefficients are negative, showing a stabilizing effect on marriages, although all are statistically insignificant. Seven of the 13 are positive, showing a destabilizing effect, but only one is statistically significant: 0.88 for the 44 black families in the five-year NIT program in Denver. The pure NIT plan does not show a consistent destabilizing effect for any of the three ethnic groups.

The NIT × TR plan has a large and significant destabilizing effect on blacks. These plans have no statistically significant effects among whites or Chicanos, although the direction of the effects for whites is mainly positive. Finally, the five-year duration plans tend to be *less* destabilizing than the three-year plans in most comparisons.

Table 6 summarizes the separate experimental plans for each ethnic group, pooling the sites and durations to build up the sample size and to summarize an overall effect of each of the three experimental treatments. Of the nine experimental coefficients, only one is statistically significant, .57 for blacks in the NIT × TR plan. Of the three coefficients for the pure NIT, none is statistically significant, and one (for Chicanos) is negative. The pure NIT coefficient for blacks, .27, is large enough to cause concern, but it is not reliably estimated, and it is smaller in absolute value than the statistically insignificant negative coefficient, —.30, for Chicanos. A weighted average for the three ethnic groups, using the sample proportions of couples in each ethnic group as weights, is .10. For the relatively rare event of a marital dissolution, an effect of this magnitude has no practical significance.

Summary

The results shown in Tables 5 and 6 do not justify the conclusion that a negative income tax program, by itself, would lead to an increase in marital breakups among married couples with children. Three telling results argue against such a claim.

1. First, as shown in Table 5, the sample sizes for the cells that describe the pure NIT plan are not large enough to warrant any confidence in such a conclusion, unless the results for the different cities,

time durations, and ethnic groups were so consistent that the samples could be pooled. But the results are not consistent even with respect to sign.

2. Second, the summary estimate achieved by combining all durations and sites in Table 6 shows inconsistent signs and an overall small quantitative effect (.10) for the pure NIT treatment.

3. Third, the results have not been adjusted for attrition bias or for reconciliations. Attrition bias is, of course, unknown, and it is merely on the basis of prior theorizing that the adjustments suggested earlier diminished the dissolution rate among experimentals relative to controls. If the reader agrees that an adjustment is called for, perhaps a summary estimate would entail multiplying all the positive NIT coefficients by .95 and all the negative coefficients by 1.05. Reconciliations are observable during the course of the experiment, and although they have not been used in this paper, the findings of Groeneveld, Hannan, and Tuma, which we have corroborated, show that reconciliations are more prevalent among the experimental families. This indicates that a measure based on the fraction of time that the parents are separated is likely to show less instability than did the rate of first dissolutions, which Groeneveld, Hannan, and Tuma and this study have emphasized.

Several qualifications must be noted about these conclusions regarding the negative income tax and marital stability. One, which is probably not serious, is that the reanalysis has not examined the paradoxical result of Groeneveld, Hannan, and Tuma whereby the least generous negative income tax plans had the largest destabilizing effect, and the most generous plans the least destabilizing effect. As stated above, it is difficult to believe that the sample sizes justify these conclusions. Second, no explanation emerges for the significant destabilizing results for the combined negative income tax-training treatment. The training plan, by itself, had an even smaller destabilizing effect than did the pure negative income tax, on average and across all ethnic groups. So it is not plausible to portray the training program as the villain in promoting marital dissolutions. The destabilizing effect from the treatment that combined a negative income tax and training program, particularly among black families, remains not well explained.

Also unresolved is the issue of conflicting biases in short-duration experiments. Are the experimental outcomes exaggerated, via the intertemporal substitution effect? Or are they understated, via the lesser present value of the incentives? This issue is particularly interesting because Groeneveld, Hannan, and Tuma had emphasized that the dissolution effect was understated by a short-duration experiment. Their evidence was their report of a stronger destabilizing effect of the five-year plans, and their claim was that a permanent plan would have even larger destabilizing effects than the five-year plan. The results in Table 5 appear to refute these claims. The tendency for three-year plans to show larger annual rates of dissolution than the five-year plans is consistent with intertemporal substitution playing a significant role.

One obstacle to further analysis of this issue is the fact that the five-year controls were nonrandomly selected from among the control groups. The 20-year plans do not offer much help on this question. Overall, these groups had lower average annual dissolution rates over the years when they were assigned, which were years three through seven. However, they also were nonrandomly selected. Both the five-year controls and the 20-year groups demonstrated the trait of "stability" by virtue of their not having dropped out during the first several years of the experiment. There was no practical (or statistically significant) difference between the 20-year treatment and control groups (results not shown), but the sample sizes were small.

What explains the contrast between the large and dramatic destabilizing results of the earlier analysis compared to the smaller and inconsistent patterns shown in Tables 5 and 6? The analysis of this question is incomplete, but all of the following appear to contribute to the new mild results:

1. Separating the NIT plan from the NIT × TR plan;
2. Eliminating couples without children from the analysis;
3. Including the couples in the 20-year plans during the years in which these plans were in effect;
4. Permitting the 20-year couples to be part of their originally assigned plans during the years when the 20-year plan was not in effect;[40]
5. Including information on marital dissolutions even if they were recorded after the date of an attrition report.

The last item refers to the apparent decision of Groeneveld, Hannan, and Tuma to record the couple as having dropped out but not as having dissolved their marriage, if attrition was reported first. Our procedure helps in a small way to correct for the alleged attrition bias. There are more dropouts among controls, and if dropouts have high marital dissolution rates, the post-attrition information helps correct for the bias.

Probably the greatest difference between their conclusions and those of this study is that they emphasized results from the first three years of the experiment including the five-year negative income tax plans. It turns out that the results for the full five years of the experiment are less adverse regarding the effect of a negative income tax on marital stability. Also, the large impact of the five-year plans they report during the first three years are dissipated when the separate plans and extra years of the experiment are included.

The prevalence of reconciliations among the sample, particularly among the experimentals, may provide a clue to the high volume of dissolution and may suggest a way in which a negative income tax plan might deal with dissolutions. Consider that the families in the Seattle-Denver plans were eligible to receive a monthly payment if their incomes were sufficiently low. Surely they would realize that a departure by a spouse with earnings, particularly the husband, would lead to a quick and sharp increase in their monthly payment. The temptation to report frequent dissolutions, along with frequent reconciliations, may be strong on the part of a small percentage of the families. Only a few dissolutions are required to make a substantial difference in the rate, when the sample sizes are small and the rates are as low as 6 percent or less per year. AFDC might provide larger benefits to "permanent" dissolutions, but, as Groeneveld, Hannan and Tuma have suggested, the fixed costs of "going on" AFDC may dissuade mothers from doing so if the separation is believed to be temporary. Perhaps a negative income tax requires a longer waiting period before higher payments are made. Obviously, more than speculation is needed to determine if the phenomenon of "temporary" dissolutions explains the high dissolution rate among black couples covered by the NIT × TR plans. Our future work will examine this issue.

The basic finding is, however, not about reconciliations. Rather, the pure negative income tax plan had neither a practical nor a statistically significant destabilizing effect on the marriages of already married couples with children.

[1]Only two among many papers by the Seattle-Denver research staff will be cited at this point. The first published article, which was especially important for being first, was Michael T. Hannan, Nancy B. Tuma, and Lyle P. Groeneveld, "Income and Marital Events: Evidence from an Income Maintenance Experiment," *American Journal of Sociology, 82,* 1977, pp. 1186-1211. The final version of their findings is "Marital Stability," in *Final Report of the Seattle-Denver Income Maintenance Experiment, Volume 1, Design and Results,* SRI International, May 1983, Part V, pp. 257-383. Volume 1 will be cited hereafter as *Final Report.*

[2]Groeneveld, Hannan, and Tuma review much of the literature up to 1980 in *Final Report,* pp. 264-266. See also David Ellwood and Mary Jo Bane, "The Impact of AFDC on Family Structure and Living Arrangements," Report to U. S. Department of Health and Human Services, 1984.

[3]Andrew J. Cherlin, *Marriage, Divorce, and Remarriage,* Cambridge, Mass.: Harvard University Press, 1981, pp. 10-11, 23.

[4]One qualification is AFDC-UP, with UP standing for "unemployed parent," an optional program offering AFDC to poor married couples whose principal breadwinner is unemployed. Now adopted by half the states, the program nevertheless has a very small number of couples participating.

[5]See the arguments and citations for a positive effect of income on marital stability in Groeneveld, Hannan, and Tuma, *Final Report,* pp. 261-64.

[6]John Bishop, "Jobs, Cash Transfers, and Marital Instability: A Review and Synthesis of the Evidence," *Journal of Human Resources,* Summer 1980.

[7]Negative income tax plans of roughly the same level of generosity can differ in their income guarantees and benefit-reduction rates, but I will not discuss the differential effects on marital stability of these sorts of variations. The trade-off between guarantees and the benefit-reduction rates was not an important issue in the analysis of marital dissolutions in the negative income tax experiments.

[8]See the interesting exchange of questions and responses on this issue in a Senate hearing on welfare reform that is reported in Gilbert Y. Steiner, *The Futility of Family Policy,* Washington, D. C.: The Brookings Institution, 1982, pp. 101-102.

[9]Groeneveld, Hannan, and Tuma cite several studies that support this argument in their research review. See *Final Report,* pp. 265, 270. See also Ellwood and Bane, "The Impact of AFDC," 1984.

[10]I am grateful to James Albrecht, for aiding my consideration of this issue. See his "Hare [sic] Today, Gone Tomorrow: Divorce, Unemployment, and Other Sorry States," in K. Lang and J. Leonard, eds., *Unemployment and the Structure of Labor Markets,* London: Basil Blackwell, 1986.

[11]For further discussion of some of the features of the Seattle-Denver experiment that created incentives for creating new family units, see Gary Christophersen, *Final Report of the Seattle-Denver Income Maintenance Experiment, Volume 2,* Mathematica Policy Research, May 1983, especially pp. 37-51.

[12]*Final Report,* 266-269.

[13]Arlene Waksberg, "Overview of Master File System with Particular Attention to the Operational Flow of Family Composition Data," p. 24. This was originally published by SRI in January 1979 and is reprinted in the documentation for the Seattle-Denver data tapes provided by the National Archives. Waksberg noted that obtaining "Affidavits of Separation" was "done in a nonrigorous fashion" (p.24).

[14]The description of the training-and-counseling treatments used in the Seattle-Denver experiment does suggest their individuality along several dimensions. See Katherine P. Dickinson and Richard W. West, "Impacts of Counseling and Education Subsidy Programs," *Final Report,* especially pp. 201-216.

[15]*Final Report,* pp. 291-292.

[16]In the Seattle-Denver experiment a minimum monthly payment of $20 was paid to experimental families who filed their monthly reporting forms. Smaller payments were made to a subset of control families who were asked to file reports. These payments undoubtedly kept attrition lower than it otherwise would have been. See Christophersen, pp. 65-68.

[17]For the evidence supporting these generalizations about attrition, see Robert G. Spiegelman, "History and Design," *Final Report,* pp. 30-32.

[18]For references to personal problems, participation in public welfare programs, geographic mobility, and marital dissolution in connection with attrition, see David N. Kershaw and Jerilyn Fair, *The New Jersey Income-Maintenance Experiment, Vol. I*, New York: Academic Press, 1976, pp. 119-127.

[19]On page 239 of the microfiche description of the Seattle-Denver experiment that is provided by the National Archives we are told only that: "Later, a sample of the control families was selected to be interviewed for the same length of time as 5-year financials [5-year NIT experimentals]."

[20]See Philip K. Robins and Gary L. Steiger, "An Analysis of the Labor Supply Response of Twenty-Year Families in the Denver Income Maintenance Experiment," SRI unpublished paper, April 1980.

[21]See the longer version of this paper, available as a Discussion Paper from the Institute of Research on Poverty, University of Wisconsin, Madison, Wisconsin. This will hereafter be cited as Cain, "Discussion Paper."

[22]For a discussion of these reporting differences and the judgment that they led to a slight bias toward more reporting of splits by experimental families than by control families, see Waksberg, "Overview of Master File System."

[23]Groeneveld, Hannan and Tuma, *Final Report*, p. 357. On page 310 the authors suggest that "reasonable adjustments for attrition bias are on the order of 10 percent for blacks and 5 percent for whites." Also, the dissolution effect they report for all couples is about 5 percent higher than that for couples with children. Therefore, the researchers' estimates of 58 and 51 percent reported above correspond in their other reported results to estimates of 64 and 56 percent.

[24]Hannan, Tuma and Groeneveld, "Income and Marital Events," 1977, p. 120.

[25]Tuma, Hannan and Groeneveld, "Dynamic Analysis of Event Histories," *American Journal of Sociology, 84*, January 1979, pp. 835-836; and Groeneveld, Hannan and Tuma, "The Effects of Negative Income Tax Programs on Marital Dissolution," *Journal of Human Resources, 14*, Fall 1980, pp. 664-665.

[26]Isabel V. Sawhill, George E. Peabody, Carol A. Jones, Steven B. Caldwell, "Income Transfers and Family Structure," Washington, D.C.: The Urban Institute, September 1975.

[27]For the evidence and citations for these claims, see Cain, "Discussion Paper."

[28]Groeneveld, Hannan, and Tuma in particular expressed skepticism that the AFDC system had an important destabilizing effect on marriage. See *Final Report*, p. 266.

[29]For further discussion of Groeneveld, Hannan, and Tuma's rather complicated explanation of their findings regarding the different levels of negative income tax plans, see Cain, "Discussion Paper."

[30]*Final Report*, pp. 358-362; "Income and Marital Events," 1977, pp. 1208-1209.

[31]Steiner also questioned the "certainty" hypothesis, but it is not clear that he was referring to the short run of immediate payments. See Steiner, 1982, p. 109. On the other hand, if Groeneveld, Hannan, and Tuma claim that certainty affects the steady state dissolution rate in the short run, they must argue that the temporary wait for AFDC benefits is sufficient to permanently dissuade the mother from her intended "permanent" separation or divorce. Would the woman choose a "permanent" divorce if she can receive negative income tax payments for, say, three months but not so choose if she has to wait three months for AFDC benefits?

[32]Bishop argues for a stigma effect of transfer payments that destabilizes marriages, but his hypothesis is nearly the opposite of that of Groeneveld, Hannan, and Tuma. In Bishop's view, negative income tax payments stigmatize the husband, demeaning his role as a provider, and in this way promote marital breakups. See Bishop, "Jobs, Cash Transfers, and Marital Instability," 1980. In contrast, the Seattle-Denver researchers argue that because negative income tax payments have relatively little stigma, they will be chosen by a divorced mother as a source of income support that she has shunned when it is available through AFDC.

[33]Christophersen, *Final Report, Vol. 2*, pp. 10-12.

[34]We use the data for the same couples as were used by Groeneveld, Hannan, and Tuma, except that we restricted our analysis to couples with dependent children (under age 21) at the beginning of the experiment, and we discarded a few cases in which either spouse died. Groeneveld, Hannan, and Tuma had discarded cases in which the wife died. (I use the plural pronoun in discussing the reanalysis to acknowledge the contribution of Douglas Wissoker.) Although an analysis of a related outcome that measures the time

when children are without both parents as a result of a marital dissolution is underway, these results are not presented here. This latter outcome is based on the information on reconciliations, which will be only briefly referred to in this paper.

[35]The level of normal earnings, in seven categories, is defined as "expected family income for the year prior to the start of the experiment, and was derived from preenrollment interview data." Christophersen, *Final Report*, p. 61.

[36]Perhaps surprisingly, the (NIT × TR)-Control difference in the dissolution rate for the five-year group of blacks is not statistically significant at conventional levels, even though the NIT × TR rate, .131, is 70 percent higher than the Control rate, .077. The P-value for the two-sided test of significance is .155. The numbers of observations used in these tests of significance are derived from the person-years of record, which are about three times as large as the numbers of couples. Thus, the levels of significance may be overstated. For example, the marital records for 10 couples for one year should convey more information than the record of one couple for 10 years. If this view is correct, the criterion for judging a difference to be statistically significant should be more stringent than usual.

[37]More precisely, the multiplier equals *e* raised to the power of the coefficient. A coefficient of .10, for example, implies that the group's dissolution rate is 1.105 times as large as the control group's dissolution rate ($e^{010} = 1.105$). A coefficient as large as .76, however, implies a multiplier of 2.14, showing a 114 percent increase in the group's dissolution rate compared to the control group. A coefficient of -.12 indicates a multiplier of .887—about a 11 percent reduction in the group's dissolution rate compared to the control group.

[38]Paul D. Allison, "Discrete-time Methods for the Analysis of Event Histories," in S. Leinhardt, ed., *Sociological Methodology*, San Francisco: Jossey-Bass, 1982, pp. 61-98.

[39]In Table 7 Groeneveld, Hannan, and Tuma show four variables for the training programs, but they combine the NIT and (NIT × TR) programs, distinguishing only the three- and five-year durations by an additive three-year dummy variable. Their specification is approximately equivalent to one in which all nine NIT and (NIT × TR) variables in Table 5 are combined, which becomes equivalent to the Groeneveld, Hannan, and Tuma "NIT" variable, and in which all three-year NIT and three-year (NIT × TR) variables in Table 5 are combined, which becomes equivalent to the Groeneveld, Hannan, and Tuma "NIT, 3yr" variable.

[40]At least I believe this is a change from the Groeneveld, Hannan, and Tuma procedure; they state: ". . . we omitted the marital histories of the 20-year families after they were assigned to the 20-year treatment. Their marital histories prior to that time (about 2 years after enrollment) are included. In our analyses all experimental families who become 20-year families are classified as 5-year experimental families until the length of treatments is changed." *Final Report*, p. 287, footnote 1. However, Robins and Steiger, 1980, had claimed that the experimental families who became 20-year families were all originally assigned to the three-year experimental plan.

Discussion

David T. Ellwood

In reading Glen Cain's paper, I was reminded of Harry Truman's expressed desire for a one-handed economist. Cain has done a careful job of discussing all the "one hands" and "other hands" that can contaminate an experiment of this sort when looking at marital dissolution. And he shows us just how unstable the results of the Seattle-Denver income maintenance experiments really are. Yet in reading this paper one is left with the fundamental question: what should we believe about a negative income tax and marital stability? In the end I certainly come away convinced by Cain's assertion that the evidence that a negative income tax is strongly destabilizing is not decisive; but I cannot fully endorse the impression of Cain's last paragraph that the experiments showed neither a practical nor a statistically significant destabilizing effect. Rather I'd say the evidence is just too thin to draw firm conclusions.

Three questions are paramount as we evaluate the possible impact that a negative income tax might have. First, should we have expected the negative income tax to be stabilizing or destabilizing for marriage? Second, what, if anything, do the experimental results show? And finally, how likely is it that the experimental results are a good reflection of what would actually occur if a "permanent" nationwide negative income tax were adopted?

What Should We Have Expected?

I was not a participant or observer during much of the period when

*Associate Professor of Public Policy, John F. Kennedy School of Government, Harvard University.

the negative income tax was being proposed and debated. But my impression from my reading and discussions was that proponents generally expected the program would be stabilizing. Personally, I think the expectation should have been that the plans, at least as implemented in the experiments, would be destabilizing.

An economic model of divorce or separation would suggest that currently married couples compare the net benefits of being married to the net benefits of being apart. Compared to a situation where there are no benefits available to anyone, a negative income tax has an ambiguous impact in theory. It provides added income to both poor intact families and poor separated ones. But in practice, the negative income tax would almost certainly be more destabilizing than doing nothing.

The financial position of intact families is very different from the position of separated ones. Two-parent families are rarely poor, and when they are, their poverty tends to be short-lived. Single-parent families are typically poor, and the poverty often lasts much longer. Thus the expected benefit to single-parent families is far greater than that for two-parent families. Of course a lack of income may be a destabilizing factor in some divorces or separations, but for the most part lack of money is likely to be a far greater problem for the split family than for the intact one. Thus even though a negative income tax appears to be neutral, it is in fact a far greater subsidy to single-parent families than to two-parent families.

Of course the proper comparison is not between the negative income tax and nothing, it is between the negative income tax and the present system. If the effect of the negative income tax was to leave effective benefits for single-parent families unchanged and to increase the economic benefits only to two-parent families, the program ought to be mildly stabilizing. But the bulk of the tested programs offered benefits far more generous than those of the existing AFDC system. Moreover, the program provided far more information on available benefits and options than would generally be known among the general public with respect to the AFDC program. Thus, although these negative income tax programs could have been stabilizing, I think it was reasonable to expect they would have the opposite effect.

What Do the Experiments Show?

Cain is very effective in showing that the results are extraordinarily confusing and unstable. I've spent two days and nights, poring over Cain's detailed numbers looking for patterns, trying to pull out what message there is. In the end, I come away mostly frustrated, unable to say anything but the most equivocal statements.

The reason really is quite simple. The sample sizes are very small and divorce or separation is a relatively rare event. The entire control group of originally intact families in Seattle numbers 263, of which 31 split up and 53 dropped out of the experiment. Breaking these even into racial groups leaves one with almost no sample. Any further breaks leave almost nothing to examine. And attrition seems quite worrisome. As Cain makes quite clear, there are very good reasons to expect far less attrition among experimental families that split apart (since they benefit more from the negative income tax) than among those that remain stable.

What makes matters worse is that the experiments included not only several sites and racial groupings but also enormous variation in the treatments received. Participants received dramatically different levels of benefits. And some experimentals were offered a variety of training and counseling programs in addition to the negative income tax benefits or instead of them. With such thin data, it is almost impossible to disentangle any of the independent effects of one program or another.

Cain argues that we ought to look mostly at the groups that received a "pure" negative income tax with no training or counseling. This is one of the few parts of the paper I found quite unconvincing. Separating the "pure" negative income tax groups from the others thins an already thin sample. Cain finds no evidence that the training and counseling programs alone have much separate effect. And he cannot offer much a priori reasoning as to why we should expect an interactive training/negative income tax impact. The main reason for separately estimating a "pure NIT" effect and an "NIT/training" effect appears to be that the impacts seem larger for the latter group. But with such thin data, surely there are many divisions that would also show highly differential impacts.

I can surely understand why one would want to include separate treatment variables for the negative income tax and training, but I do not see why we should so severely limit our sample sizes in order to allow for an interactive effect of the NIT/training treatment combination. To my knowledge none of the labor supply models employed such a methodology, even though one could argue more directly for a possibly joint effect in that situation. Nor can we say, if a negative income tax plan were actually implemented, whether or not it would be accompanied by a training-like component.

Normally the way we deal with small samples is to pool. But we do so at our peril, of course. These data show little consistency across sites and treatments. Cain's "pure NIT" was stabilizing for blacks in Seattle but strongly destabilizing for them in Denver. Chicanos, on the other hand, were stabilized in Denver. Results for whites were similarly perplexing. As a result the standard errors of all the estimates were extremely high. One clearly cannot infer much from these data. The samples are simply too small and unstable to say anything definitive.

Yet the percentage point estimates are troubling. The overall effect of the negative income tax was to push up family splits among whites by 18 percent, and among blacks close to 50 percent. Even if one looks only at the "pure NIT" as Cain urges, destabilizing effects in the range of 15 to 30 percent are found for whites and blacks. The confusing Chicanos showed a moderately stabilizing pattern. Even though few results are significant, I conclude that there is almost no evidence in these results to suggest the Seattle-Denver income maintenance experiments were stabilizing; however, I do not think we can say just how large the destabilizing effects were. It is worth remembering that significant impacts were not found in other experimental sites.

Would the Results Be the Same for a National Negative Income Tax?

I see no reason to infer much from these results. Cain points out that a short-duration experiment could have a smaller than actual effect because people cannot count on the support indefinitely, or a larger than actual effect if people divorce now while there is an unusually generous basis for support. I strongly favor the latter hypothesis. Divorce or separation is certainly a "threshold" event where some impetus ultimately pushes people into action. It is also an event that may have a short time horizon. Many couples see separations as temporary or exploratory. Moreover, most separated women remarry or reconcile rather quickly. Finally, I doubt many women who go on welfare after a divorce see it as anything more than a temporary bridge. Data on the AFDC program show that formerly married women have the shortest durations on welfare.

The negative income tax may have been seen as a unique moment when a transition into another living arrangement was easier. And the information and attention that experimentals received may have brought the financial options into clearer focus. Some evidence for the proposition that information could have had an impact in and of itself comes from the fact that Groeneveld, Hannan, and Tuma found large destabilizing effects even for plans where benefits were no higher than under current AFDC programs, which pay only for single parents. Theory is unambiguous in suggesting that a program that leaves benefits unchanged for single parents while providing new benefits to two-parent families should be stabilizing. The fact that such plans were destabilizing suggests that information or some other factor contaminated the results. And one possible interpretation of the apparently higher impact of the negative income tax/training counseling combination could be that these programs helped people see how the negative income tax might help them in the short run. (I still do not think this explanation is plausible enough to justify special treatment of the option, however.)

In general, then, I think we learned very little from the negative income tax experiments with respect to divorce and separation. We learned that the experiments did not stabilize families and that they may have been destabilizing. But since I have argued that we should have expected that result anyway, I'm not sure that is very valuable information. I am less skeptical than Cain about the prospects for learning something about these events through experimentation. I think the biggest problem here was that sample sizes were small. But I do believe that these events are inherently difficult to study and are likely to be severely influenced by the experimental design itself, independent of the changed incentives that may be created. Yet social scientists interested in poverty must explore these issues, perplexing and ephemeral as they may seem, for family structure changes and poverty are inextricably and increasingly related.

Discussion

*Nancy Brandon Tuma**

Glen Cain's paper has "two main messages:" (1) "that the evidence [about the effects of the negative income tax experiments on marital stability] is not decisive, or even persuasive;" and (2) "that family issues like marital stability are not well-suited to experimental research." I will comment on each.

Is the Evidence Decisive or Persuasive?

Is the evidence about the effects of negative income tax treatments on marital stability decisive or persuasive? Cain says "no" to both parts of this question. I agree with him that the evidence is not decisive, but I disagree with him about whether it is persuasive.

A decisive result is rare in any experiment, whether it tests a new drug for treating cancer or a new weapons system. At best, most experimental results turn out to be "persuasive" or "suggestive." That is, they alter one's best guess (and hypotheses for the next study), but they are almost never so definitive that a next study is unnecessary.

Our analyses (I refer to those by Groeneveld, Hannan and myself, and especially those described in our final report) convinced me that the negative income tax treatments decreased the marital stability of low-income black and white couples.[1] Cain's reanalyses, which are, in fact, very similar to various analyses included in our final report, have not altered my conclusions. Although a detailed comparison of our 125-page final report and Cain's paper is not possible here, I will summarize what I consider to be the most salient points.

*Professor of Sociology, Stanford University.

Cain presents results for analyses that differ from ours in a number of relatively minor ways, many of which had already been explored in our report and were known (from our reported results) to decrease somewhat the negative income tax's effect on the marital breakup rate of black and white couples. When all of these minor changes are put together, Cain finds a positive but statistically insignificant effect of what he calls a ''pure'' negative income tax treatment on the marital breakup rate of black and white couples.

The statistical insignificance of Cain's finding is no surprise because the power of hypotheses tests about marital breakup rates using the Seattle-Denver data is low. In order to achieve *statistical* significance, both our analyses and Cain's found that the negative income tax treatments would have to increase the marital breakup rate by roughly 40 percent for black and white couples (and by over 80 percent for the much smaller sample of Chicano couples); hence, increases in the breakup rate that are smaller than 40 percent cannot be statistically distinguished from ''no effect,'' although they may be big enough to be of considerable *social* significance.

Even if one accepts Cain's analytic decisions that act to reduce the negative income tax effects, his ''pure'' effect (see his table 6) is still positive and large enough to be noteworthy: his estimated increase in the marital breakup rate is 17 percent for whites and 31 percent for blacks. Moreover, the ''impure'' effect of combined negative income tax-training treatments is as large and positive as the ''pure'' effect for whites and much larger for blacks. Most people would not ignore the ''impure'' effect of the combined treatments, especially since the ''pure'' training effect is tiny for whites and moderate for blacks.

In addition, one may not want to accept Cain's analytic decisions for the following reasons:

(1) Cain omits childless couples because he believes that any negative income tax programs passed by Congress would exclude them from benefits. I contend that our job as social scientists is to analyze *all* of the data. We recognized that the presence of children might affect response, so we did estimate some models with separate effects for couples with and without children. We found (Groeneveld, Hannan, and Tuma 1983, table 5.8, pp. 298-99) that the negative income tax effects for couples with children were smaller than those given in our summary in the case of whites (a 36 percent increase rather than a 53 percent increase) but were about the same for blacks. That is, the negative income tax effect was in the ''40 to 60 percent'' range (a summary figure from our conclusion on which Cain focuses) for blacks, but a little less for whites.

(2) When Cain analyzes similar data using the same explanatory variables with a similar model, his estimates (table 7) are only about

80 percent as large as ours. (This applies to effects of nonexperimental variables as well as experimental treatments.) I suspect that his estimates are smaller than ours because he aggregates the data based on the date of the dissolution. (Cain aggregates to six-month intervals; we recorded events to the nearest day.) Although Monte Carlo studies are needed to say for sure, time aggregation probably biases estimates downward. Cain's decision certainly has no known scientific advantages.

(3) Due to the small sample size relative to the number of treatment and assignment variables, analysts of these data cannot cross-classify by all treatments and assignment variables. Cain chose to ignore one set of cross-classifications; we chose another. Naturally, results depend on these choices. Whose choice is better? Two differences in our choices stand out:

(a) Cain stresses a model that includes an interaction between training (actually, a mixture of three quite different treatments) and the negative income tax treatment (a grouping of 11 different financial plans). Like Cain, we estimated a model that interacted the negative income tax treatment with the training treatments, but we separated the three training treatments, which we regarded as quite different. We found that the set of interactions was not significant for blacks but was significant for whites (Groeneveld, Hannan, and Tuma 1983, table 5.B.2, pp. 371-72). We were skeptical about these results, however, because the pattern of effects for various treatments was unsystematic. We thought that the data were cross-classified so much that chance variations due to small cell sizes swamped any trends. Omitting the negative income tax-training interactions gives what we consider to be a clearer view of the overall effects on the negative income tax treatments.

(b) Cain handles plan length (length of treatment) differently than we did. In our view, having a five-year plan rather than a three-year plan is analogous to giving a drug to cancer patients in two strengths, the first more potent than the second. In this parallel situation, analysts do not regard the two treatments as entirely unrelated. Rather, they test whether the effects of the two doses differ. If patients given the stronger dose respond to the drug significantly, and patients given the weaker dose have a similar but smaller and insignificant response (essentially what we found), most analysts conclude that the drug *does* have an effect, but that one dose was too weak for its effect to be detected with the data available. This reasoning led us to stress the effects of the five-year plan.[2]

Cain's approach is quite different. In his table 5 he interacts plan length with site and his three treatment components: training, negative income tax, and negative income tax-training. I do not see any scientific reason for his approach here, but I would predict that spreading the treatment effects across 13 treatment variables is extremely unlikely to yield systematic or significant effects. In table 6 he omits not only the site and plan length interactions, but also a main effect for plan length. This is like treating weak and strong doses of a drug as equivalent. I do not see any scientific grounds for this.

Finally, there is a piece of evidence from our final report that Cain does not mention and that helps convince me that the negative income tax treatments did increase marital breakup rates. Namely, we also analyzed pooled data from the Seattle, Denver, and New Jersey experiments, which increases the overall sample size substantially. The larger sample increases the power of tests and greatly reduces the standard errors of estimated effects. These analyses gave estimates of significant, 25 to 35 percent increases in the marital breakup rates of white, black, and Hispanic couples (Groeneveld, Hannan, and Tuma 1983, table 5.11, p. 303). Further study is needed because we did find some important variations with site.[3] Still, the evidence from the pooled experimental data is persuasive that the negative income tax treatments tended to have *some* positive effect on marital breakup rates of low-income couples in diverse settings.

In comparing Cain's analyses and ours above, I have stressed differences that in principle can be evaluated objectively. Another difference may arise from our disciplinary perspectives. As an economist, Cain stresses monetary differences between the negative income tax treatments and welfare programs like aid to families with dependent children (AFDC). As sociologists, we consider nonmonetary as well as monetary differences in these programs. Much of our final report was devoted to analyses that tried to understand why a negative income tax program that was financially similar to AFDC increased marital breakup rates. Indeed, we thought this "message" was as important as numerical estimates of an overall negative income tax effect on marital stability, which is what Cain emphasizes. Since Cain "assumes away" this part of our message, I will restate it.

We argued that administration of AFDC and of the experimental negative income tax programs differed in several key ways that could cause differential response to the same monetary benefits. (1) Knowledge of benefits and rules is likely to be lower for AFDC than for the negative income tax programs, which were carefully explained initially and again a year later. (2) The costs in time, effort, and social embarrassment of getting benefits is greater with AFDC than with the negative income tax treatments. The latter, for example, required only a monthly mailed report of income and family composition, and the same report was to be sent whether or not a breakup occurred. (3) Promptness and the short-run certainty of receiving benefits after a breakup were greater in the negative income tax program than with AFDC, again because no special action on the part of a recipient was required after a breakup in order for benefits to begin or to be increased (if the couple was already receiving benefits). Unfortunately, with the data from the Seattle-Denver income maintenance experiment, one cannot assess the *relative* importance of these differences between AFDC and the experimental negative income tax programs. But we think that together they account for the relatively large increase in marital breakup rates under negative income tax treatments financially similar to AFDC.

Three final points about nonmonetary differences between AFDC and negative income tax programs deserve mention. First, although the administration of the *experimental* negative income tax programs made it easy for people to benefit from them, a *federal* program might not be administered in a similar way. Second, one could experimentally vary administrative features of a negative income tax program and study the consequences. Third, *if* administrative features are important, as we argued, and *if* the administrative features of a federal negative income tax program are different from those in the Seattle-Denver experiment, then neither our numerical estimates of negative income tax effects nor Cain's are a good basis for estimating the costs of a proposed federal program.

Are Family Issues Suited to Experimental Research?

What about Cain's other message? Should family issues be studied experimentally? Cain says "no," primarily, it appears, because he believes the cost of a well-designed experimental study would be "too high." Deciding if one agrees with Cain requires a cost-benefit analysis involving answers to three questions:

(1) What would a well-designed experimental study of family issues look like? What would it cost? No one has yet tried to *design* such a study, let alone estimate its *cost*. Thus, a very basic piece of evidence for Cain's view is missing.

(2) What would it cost to obtain the same information by other means? The most likely other source of such information would be analyses of nonexperimental data, for example, panel surveys. Not only would a good nonexperimental study of family issues be costly, but it quite possibly might be more costly than a well-designed experiment.[4]

(3) How valuable is knowledge about the relationship between social policies and family issues? Whatever the cost of a study of family issues, some people may think it is "too high" simply because they don't value the information it produces.

Since Cain has not yet given serious answers to (1) or (2), let alone said how much he thinks such information is worth to our society, I am totally unconvinced by his claim that the costs of a properly designed experimental study of family issues would be "too high."

I am convinced, however, that someone needs to think hard about how to design good experimental and nonexperimental studies of family issues, so that debate about the value of such studies can move from the level of rough and ready speculation to one with a sound scientific basis. And, while I am persuaded that the experimental negative income tax programs tended to decrease marital stability of low-income couples, I also think estimates of the magnitude of these effects (both ours and Cain's) are not sufficiently precise for policy planning. Moreover, the negative income tax experiments definitely did not give adequate information on the role of *nonmonetary features* of the treatments. If there is another set of negative income tax experiments someday, I hope that they will be designed not only to obtain more precise estimates of effects of plan generosity but also to study this important issue.

[1]Like Cain, we concluded that the negative income tax treatments did not alter the marital breakup rate of Chicanos couples. However, we did find that they markedly decreased the marital *formation* rate of unmarried Chicana women with children, and this is an alternative way of decreasing marital stability. So, while I think the evidence shows that the negative income tax treatments decreased the marital stability of Chicanos, this cannot be detected from analyses of data on marital breakups from the Seattle-Denver income maintenance experiment. See Groenveld, Hannan, and Tuma 1983.

[2]Cain suggests that his results may differ from ours partly due to different handling of those assigned to the 20-year plan. As far as I can tell from his paper, Cain treated them exactly the same as we did. In any case, I'm skeptical that somewhat different handling of fewer than 10 percent of the sample would cause appreciable differences, especially since our analyses focused on the first 36 months of data and the 20-year plan only began after about 30 months.

[3]The effects of the negative income tax treatments on marital breakup rates were somewhat smaller for whites in New Jersey than for whites in Seattle and Denver. Contrarily, the effects were much larger for Hispanics in New Jersey (mainly Puerto Ricans) than for Hispanics in Denver (mainly Chicanos). The variation with site could arise because of cultural, ethnic, or religious differences in the populations in the three sites. They could also be partly due to differences in state programs of aid to families with dependent children; these differences cause control group comparisons to differ even if the negative income tax plans are the same. However, the negative income tax treatments in New Jersey also differed in a number of ways from those in Seattle and Denver, so this is yet another possible reason for differences across sites. Still other reasons for site differences can be suggested.

[4]Since available nonexperimental data on family issues and income are still very inadequate, a good nonexperimental study would almost certainly involve costs of data collection as well as analysis. And, since the costs of the negative income tax experiments came disproportionately from data collection and analyses—not from administration of treatments (see Zellner and Rossi 1986)—a nonexperimental study might not cost much less than an experimental study with the same number of cases. Moreover, sample sizes must usually be much larger in a nonexperimental study than in an experimental study, in order to estimate effects with equal precision. As a result, a good nonexperimental study of family issues could be more costly than a well-designed experiment. This ignores likely biases in nonexperimental studies, which are even harder to handle than the two sources of bias that Cain associates with an experiment.

References

Cain, Glen G. "The Issues of Marital Stability and Family Composition and the Income Maintenance Experiments," 1986, this volume.

Groeneveld, Lyle P., Michael T. Hannan, and Nancy Brandon Tuma. "Marital Stability" in *Final Report of the Seattle-Denver Income Maintenance Experiment, Volume 1: Design and Results,* Washington, DC: U.S. Government Printing Office, 1983.

Zellner, Arnold and Peter E. Rossi. "Evaluating the Methodology of Social Experiments," 1986, this volume.

Non-Labor-Supply Responses to the Income Maintenance Experiments

*Eric A. Hanushek**

The concept of a negative income tax has been actively discussed and promoted, at least by economists, for over two decades. High on the list of motivations for this are the inefficiencies and inequities of patchwork welfare programs that make arbitrary distinctions among potential recipients and concentrate on specific consumption items. The possibility of extremely high marginal tax rates on benefits, resulting in part from enrollment in multiple programs, also has contributed to interest in a negative income tax. The majority of the policy discussion has focused on the labor supply effects, which have so much potential influence not only on program costs but also on public perceptions of the welfare system. The centerpiece of the analysis from the various income maintenance experiments has always been the statistical manipulation of labor supply data. Invariably, however, residual analyses, typically described as "non-labor-supply results," are also included, and a portion of these results that do not involve the structure of the family forms the subject of this paper.

Since the focus of the experiments was so confined to labor force issues, design features in the other areas were not given the same degree of attention. At the same time, the detailed data have provided a good base for a variety of analyses, heightening the benefits of the experiments per se. The tag-on nature of much of this research is understandable. First, in a wide variety of possible non-labor-force

*Professor of Economics, University of Rochester. The author is indebted to Edward Gramlich, Bruce Jacobs, Charles Metcalf, Walter Oi, Charles Phelps, John Quigley, and Susan Silverman for valuable comments and suggestions.

effects there is no clear idea of what might be desirable. Underlying much of the negative income tax philosophy has been the notion that categorical, restricted aid programs tend to be inefficient because they do not recognize the specific preferences of the recipients. A corollary of this is that we do not have good notions of what kinds of spending behavior are most desirable. Second, a wide variety of areas have no real benchmark. We know relatively little about specific consumption patterns and how they vary across households, and, most specifically, about the overall pattern of spending by the poor as compared to the nonpoor. These issues interact with the interpretative problems that naturally arise in complex experiments: problems of sample selection, limited time horizons, imperfect experimental design, and data collection and measurement difficulties clearly affect the ability to generalize from the specific results.

For expositional purposes, if not substantive ones, it is convenient to divide the analyses into "consumption" and "investment" outcomes.[1] The reason for this division is clear. We have few firm opinions about the desirability of any consumption bundles chosen, and even a complete understanding of the determinants of consumption decisions is unlikely to have much influence on policies. On the other hand, investment-type activities are presumed good since they might lead to longer-run beneficial effects in the alleviation of poverty.

These categories clearly have fuzzy boundaries. Consumption by children, for example, might in fact be viewed as an investment, since better nutrition or housing may lead to long-run improvement in their welfare. Indeed any expenditures on children are frequently lumped into the "investment" category, because they tend to facilitate the development and learning of the next generation. For the most part, however, precision in the categorization is not all that necessary.

The data base for this paper is the vast amount of research engendered by the experiments and conducted by both the principal contractors and others. Simply extracting significant coefficients where they are found would clearly be misleading, however; doing so would obscure the volumes of regression estimates produced and would not highlight the issues most central to program policy considerations. Further, the distinctly different approaches to the same problem make quantitative comparisons virtually impossible in many areas. (See Hollister 1978.) This review will consequently be restricted to a smaller number of key areas. The emphasis is on identifying common findings that might be generalized. Whenever possible, the review refers to the books and journal articles coming from the experiments, on the grounds that these are generally more accessible than unpublished research papers or even the final reports on the experiments.

What Do We Expect To Observe?

Several factors affect what we would expect to observe in terms of consumption and investment responses to a negative income tax. First, we generally feel more confident about understanding behavioral responses to differences in permanent income than responses to transitory changes. With an increase in permanent income, people are expected to increase their overall consumption standards. With an increase in transitory income, the responses are less predictable and less interpretable because transitory changes are not necessarily shifts in the budget constraint. Therefore, when families receive income supplements under a negative income tax, their responses would probably vary depending upon whether or not they considered this to be permanent or transitory income. This is in large part a question about how individuals respond in the experimental setting.

A second issue is the dynamics of consumption. When adjustment costs are significant, individuals may not adjust immediately to changed circumstances. The clearest place to see this is in housing consumption. The Housing Allowance Demand Experiment suggests that only about one-third of the full adjustment occurs in the first year (Hanushek and Quigley 1979). Such slow adjustments cause severe problems in analyzing the short data series from the negative income tax experiments, because substantial portions of the complete adjustment cannot be observed over the course of the experiment. Clearly such lags can be explicitly incorporated in any analysis.[2] But the experimental analyses, particularly outside of the area of labor supply, rarely have pursued these issues.

Both of these issues suggest that the short-run effects observed from the negative income tax experiments might be poor estimates of the longer-run effects that would be observed in the steady state under a permanent and fully operational negative income tax. In some cases, the direction of bias is clear; for example, with investment in quantity of schooling, discussed below. In other cases, such as marital dissolution or fertility, the issue is less definite.

Finally, through the experiments, analysts have discussed the possible biases introduced by such things as sample design or attrition. When the findings depend upon estimates of mean differences between experimentals and controls, the estimates are a function of the precise sample employed. With nonrandom samples, estimated experimental effects alone are insufficient for policy purposes. To generalize these findings to a larger universe—one that differs in systematic ways from the experimental families—one must understand more fully the underlying structure of these behavioral effects. This is a difficult task in general, and the sample characteristics become more important. The problems are undoubtedly less when any generalizations are based upon more fundamental behavioral estimates, such as estimates of income elasticities or the effects of family size on educational outcomes. The problems do not, however, go away. In these latter cases, which are conditional upon the observed family circumstances, it is still necessary to ascertain whether or not the probabilities of being in the sample are related to the investment and consumption decisions being analyzed.

Consumption Patterns

Two issues arise when considering the impact of a negative income tax on consumption patterns. First, a concern about the consumption levels of the poor motivates many to support transfer programs, but little is known about their actual consumption patterns. Large gaps remain in our knowledge about the consumption choices of the poor and the resulting patterns of expenditures and well-being across families. Second, and more importantly, no criteria exist to rank alternative outcomes. If, for example, we observe that families under a negative income tax purchase more clothing, what should we think about that? Is that good or bad? In a few areas we at least bring some preconceptions to the problem, but in most we have nothing to go on.

One obvious motivation for the study of consumption aspects of income maintenance experiments is the suspicion by some that subsidies will be used for frivolous expenditures—color TVs and fancy cars—rather than for the necessities of life.[3] The measurement of consumption in the experiments is very difficult, and none of the survey efforts appeared to do very well on this score. Nevertheless, the results suggest no general increases in frivolous or outlandish expenditures. Indeed, expenditures induced by experimental treatments follow (at least in aggregate) the same patterns observed from nonexperimental income. In other words, for most expenditure categories such as food, clothing, health expenditures, and so forth, the results show nothing startling or unexpected.

The area where increased consumption by the poor is most commonly recognized as a positive outcome is housing. Housing has always received special attention in public policy matters. This may reflect a general view that housing is a basic necessity and that a just society would provide safe, decent, and affordable housing for all of its citizens. It may also simply be a reflection of the preferences of the donors—that is, dilapidated housing is offensive to others and something should be done to eliminate blight in housing markets. Part of this is an externality argument that poor housing conditions lower the property values of others in the community. Part of it is simply a desire not to be confronted by poverty that is inescapably obvious, as is the case with slum housing.

Moreover, for any quality level of housing, homeownership is frequently rated as superior to rental. Owners are more likely than tenants to maintain their homes, thus providing a superior stock of housing for the poor over time. Moreover, home purchase provides a very common way to accumulate wealth, which might give the poor the means to escape poverty.

In fact, housing policies and negative income tax proposals have often been considered together, including the conduct of parallel experiments. The housing experiments, conducted with many variants on the basic formula for the Housing Allowance Demand Experiment in Phoenix and Pittsburgh and with a saturation design in the Housing Allowance Supply Experiments in Green Bay and South Bend, provide a useful benchmark for the housing consumption results in the negative income tax experiments. The most common form of housing allowance considered is a negative income tax subsidy formula, with the guarantee and tax rate scaled to reflect housing costs and the fact that housing represents about 25 percent of total expenditures. Housing quality and rent standards are added to this subsidy formula, generally as eligibility criteria. The housing standards ensure that only people living in "suitable" housing receive the subsidies. (See Bradbury and Downs 1981 for a thorough review of these experiments.)[4]

The results of the combined studies in terms of expenditures on rental housing were surprising at the time. Before the experiments, it was commonly presumed that income elasticities for housing were approximately one. The negative income tax experiments and the housing allowance experiments consistently indicate that income elasticities of housing are relatively low for the poor: a 10 percent increase in permanent income, from a subsidy or from another source, implies an increase in housing expenditures of 2 to 3 percent in the short run and around 5 percent in the long run.[5] To be sure, it is difficult to obtain precise estimates of these elasticities because of the short-run nature of the experiments and the lags in adjustments discussed previously, but plausible adjustments do not affect the conclusion that income elasticities are considerably lower than previously believed. In other words, one might infer that the poor do not appear to view quality of housing as their most important problem, because they tend to spend only a small part of any added income on housing.

A second finding is more surprising. Analyses of data from the experiments in Gary (Kaluzny 1979) and Seattle-Denver (Ohls and Thomas 1979) indicate that the income maintenance programs tend to encourage homeownership. In fact the estimated effects appear to be quite strong. For example, at the beginning of the Gary experiment, 23 percent of the experimental households owned homes; this rose to 34 percent three years later. Of the increase, 4 to 6 percentage points appears to be a treatment effect (Kaluzny 1979).

One would expect that the temporary nature of the experiments—something not included in the housing analyses—would mute any effect on housing ownership.[6] Nevertheless, the estimated homeownership effect, which is reasonably consistent across the Gary and Seattle-Denver experiments, suggests some noticeable experimental reactions that could potentially have long-term consequences.[7] This result, however, may simply reflect "timing" effects. The addition of transitory income during the experiment might move up the time when a household has the means to make a housing purchase that it would otherwise have made sometime later. (This is similar to the finding of Dynarsky and Sheffrin (1985) on the homeownership effects of transitory income.)

The expenditure evidence from the negative income tax experiments is quite similar to that from the housing allowance analyses, the largest difference being that the housing allowance experiments obtain lower participation rates. This is almost certainly related to the necessity in many cases to move in order to take advantage of the housing programs—something eligible households might be unwilling to do in a short-run program. In the longer run, participation in an ongoing program would undoubtedly be higher than that observed in the housing experiments, but the magnitude of adjustments in housing consumption would probably stay low.

Investments in Human Capital

Investment in human capital appears very relevant for negative income tax policy. A negative income tax program operates directly on households' work incentives and rewards from market activity, which in turn affect households' investments in skills. The analytical problems surrounding the experiments are, however, quite severe because the returns to any investments in human capital will accrue over the entire lifetime and for all practical purposes will not be observed during the experimental period.

Two aspects of schooling have received attention during the experiments. The first is the extent to which a negative income tax program alters the school-work choices of youths in experimental families. The second is the effect experimental treatment has on the scholastic performance of school-aged children. Another form of human capital investment decision—entering into vocational training programs—has received less attention.

Quantity of Schooling

The influence on quantity of schooling obtained by youth is the more direct and observable investment effect during the time of the negative income tax experiments. The decision about school attendance or job entry (or neither) is clearly affected by both the costs of attending school and the subsequent returns through the working lifetime. A negative income tax subsidizes schooling by reducing the cost of not working, where forgone earnings are the most important costs of attending formal schooling.[8] What Gary Burtless (this volume) called a ''sale on leisure'' can also be interpreted as an increase in general scholarship funds. This effect on the costs of schooling will be the same in both short-run and long-run program operations.

The effects on returns are more ambiguous. A short-term experiment will not involve any important effect on returns to schooling. Longer-run effects will depend importantly on the generosity of the program and on the level of skills acquired. A very basic program might have no effect on the returns to investment if the child were above the breakeven point both before and after any marginal investment. At lower levels of investment or with more generous programs, an ongoing negative income tax program would, however, operate to lower the potential returns from an investment in schooling. Because of the potential effect on returns, experts disagree about what should be expected in terms of investment incentives with an ongoing program; (see Venti and Wise 1984; Rea 1977; and Weiss, Hall, and Dong 1980). In the case of a basic program, the experimental evidence would give a fair indication of long-run effects. In the case of lower levels of investment or more generous programs, we would expect any observed increase in school attendance for negative income tax recipients in the experiments to be an exaggerated statement of the likely ramifications of an ongoing program.

The analyses of schooling decisions have been conducted in a variety of ways. The most interesting consider, in one way or another, a trichotomous choice: work, schooling, or leisure. In each case,[9] the experiment appears to have had a positive effect on school attendance by youth in experimental families, along with a reduction in work activity. In fact, the results are strongly consistent across analyses: youth tend to increase schooling by about the same amount that they decrease labor supply, leaving leisure essentially the same as it would be without a negative income tax.

Because of the different specifications of the models, it is very difficult to summarize the quantitative impacts. Nevertheless, the estimated effects appear quite large and significant. For example, Mallar (1976) estimates the probability of completing high school for families on a "middle" negative income tax plan to be 25 to 30 percent higher, with a one-half year increase in schooling for 18- and 19-year olds during the three years of the New Jersey experiment. Venti and Wise (1984) find an 11 percent increase for youth in the Seattle-Denver experiments. They also find that increases in schooling for experimental individuals are smaller among blacks than whites and greater for females than males.

The long-run implications are, as mentioned previously, still subject to question. Nonetheless, significant increases in school attendance may well result from a negative income tax because of the substantial subsidies that arise from reducing the opportunity cost of attending school. The full implications of this would, of course, also consider the rate of return to any increases in schooling.

The Seattle-Denver experiments present an additional policy investigation. One set of experimental treatments involved counseling and subsidies directly related to education and training. All participants received free counseling, while other groups received half or full payment of tuitions and other direct costs of training.[10] Thus, in these experiments it was possible to distinguish general "income" effects related to program subsidies from direct training allowances. The idea behind these experimental treatments is clear; through training inducements, it was hoped that individuals' human capital could be augmented sufficiently to offset some of the adverse labor supply effects. (A good description of these experimental treatments can be found in Hall 1980.) The explicit training subsidy of the Seattle-Denver experiments, however, appeared to have little effect on school attendance beyond those previously noted.[11] This finding undoubtedly reflects the relatively small direct costs of schooling for most of these potential students. (Note, however, that the effects and costs of such a subsidy program might differ dramatically from the sample observations if the program were opened up to unattached youth not living with their parents.)

Scholastic Performance

The analysis of educational performance in the experiments has been conducted within the general framework of educational production functions (compare Hanushek 1986b).[12] Various output measures are related to characteristics of families, friends, and schools. Additionally, within the experiments, an independent experimental treatment effect is estimated.

Before considering any specific evidence, it is useful to review why we might expect any effects from the experiments. Previous studies of educational production have invariably found that family background is extremely important in determining the scholastic achievement of children. These studies have typically included some measures of socioeconomic status of the family as an indicator of the educational inputs in the home. But this work has for the most part not been very concerned about the details of the family effects or the underlying causal structure—things that are more important for evaluations in the negative income tax context.

The most common interpretation of the relationship between scholastic performance and socioeconomic status of the family is that socioeconomic status proxies a set of attitudes, abilities, and patterns of learning within the home. These would not be expected to change very quickly with short-term changes in economic circumstances. Thus, to the extent that the negative income tax experiments lifted the current economic situation of the family without changing these more fundamental factors, one would not expect to observe much effect on children's performance.

Nevertheless, a negative income tax might affect school performance through several routes. The most direct impacts on school performance might come from the tax's labor supply effects. Inputs of parents' time into children's learning have been a central concern of many researchers looking at the education of children. (See, for example, Leibowitz 1974, and Hill and Stafford 1974, 1980.) It is frequently asserted that inputs by the mother are most important, and, if so, this links education closely to a negative income tax, where secondary workers seem particularly sensitive to the labor supply incentive effects. The evidence, however, suggests that the relationship between mother's labor supply and children's achievement is weak (Murnane, Maynard, and Ohls 1981; Hanushek 1986a). Similarly, if a negative income tax encourages marital dissolution, the removal of one parent may well have direct educational effects.

Beyond direct time input of the parents, one would naturally look to direct improvements in the health and environment of the families and children. If a negative income tax leads to better nutrition, more effective expenditures on health, and to generally improved housing, the overall capacity of children to learn could be improved.[13] Improving housing may also involve shopping for better schools. Consumption expenditures that cut down on the time required to do household chores could also free time for parenting and educational purposes. Finally, in the more long-run category, any impacts on the number of children in the family could also filter back into educational performance. The extensive literature on family size and achievement supports the general notion that average achievement is lower in larger families.[14]

Systematic evaluations of school performance were conducted in the rural and the Gary experiments, each of which collected school data to supplement the already available household data. The Seattle-Denver experiments made a much less serious attempt at collecting data that would be useful in the analysis of educational performance. Specifically, they did not have good measures of the characteristics of the schools attended by the children.[15] The discussion here concentrates on the rural and Gary experiments.

In each case, the methodology was straightforward. Standardized test data, absences, and school grades—collected from school records—were used to measure performance. Regression equations were estimated to explain individual student variations in performance as a function of preenrollment characteristics of the families and preenrollment performance on tests for the measure of outcome considered. A variety of school characteristics were also included to account for nonrandom differences in school and classroom assignments. A dummy variable was then included to indicate experimental status.[16] The interpretation of this experimental effect is simply the average performance change of students in experimental families compared to control families.

The direct experimental evidence on any relationship between treatment and scholastic performance is mixed. For the three separate experimental groupings (Gary, rural Iowa, and rural North Carolina), the most systematic experimental effects related to test score performance in the lower grades. Children in experimental families tended to improve relative to children from other families. In higher grades and in non-test score measures of performance, no generally significant experimental effects were found, although results differed somewhat across the samples.[17] Further, the experimental effects and consistency of the findings were greater in North Carolina (Maynard and Crawford 1976) than in Gary (Maynard and Murnane 1979); experimental effects on school performance were nonexistent in Iowa. Maynard and Murnane explain the different results by the generally more deprived backgrounds of children in North Carolina, but this hypothesis is not tested directly. In the Gary analysis, time in the experiment influenced achievement gains. Children who had been in the experiment for three or four years did significantly better than children in control families or children who had not been receiving the experimental treatment for as long a period of time.

The explanations of direct experimental effects on scholastic performance emphasize parental time effects. However, as noted above, these must come from fairly subtle factors since direct testing found no relationship between labor supply of parents and children's achievement. Other evidence on schooling plus the implausibility of inducing general changes in the educational environment of the home suggest that any estimated effects of experimental treatments on educational performance should be discounted. Experimental evidence simply provides little information about the long-run, steady-state effects on scholastic achievement.

Generalizing from the Experiments

When prices and incomes vary across geographical areas, it is difficult to generalize to a national experience. The housing analyses provide the simplest example. Behavior varied significantly across the sites of the housing allowance experiments, particularly with respect to program participation. Changing conditions in the housing markets could account for the variation (compare Hanushek and Quigley 1981), but this explanation leaves some question about how to make generalizations in terms of expenditures, quality, and adjustments. Moreover, since housing represents a large fraction of a typical household's expenditures, such variation will filter through to other aspects of the consumption bundle.

Variations across sites also show up in the analysis of education. For example, the experimental effects estimated for scholastic achievement varied dramatically across sites. Maynard and Murnane (1979) hypothesize that the differences reflect the proportional differences in the amount transferred relative to initial incomes, but they do not test this directly. Venti and Wise (1984) find very different estimates of the college attendance induced by the income maintenance experiments in Denver and Seattle. In particular, Seattle youth—who are more likely to attend college in the first place—are found to react much more to the experimental treatment than Denver youth. No attempt is made to explain this difference.

Variations in behavior across regions are not easy to explain by economic theory and leave tremendous uncertainty about generalization from the experiments. Neither program costs nor participant behavior can be extrapolated easily. With the small number of experimental sites, there is no reason to presume that the sites are representative of the population or that the observed reactions in any way bound the range of behavior that would be observed in a national program. To all this must be added the previously discussed issues about limited duration experiments and time of adjustments. Certainly progress has been made on understanding some aspects of the dynamics, particularly with the Seattle-Denver variations in experimental length. But the uncertainty about results that arises from this source is difficult to eliminate.

Conclusions

The issues of the negative income tax experiments considered here are less central to the overall policy deliberations than the issues of labor supply or family composition. Moreover, in virtually every area considered here, cheaper and more direct ways exist to construct data bases and do analyses than through an experiment.[18] To the extent that major policy concerns remain about consumption or investment aspects of a negative income tax, a supplementary research program would provide more definitive estimates of behavioral reactions.

Consumption effects of a negative income tax program are difficult to observe or estimate from the experimental data. Besides the general analytical difficulties in this area, the limited duration of the experiments inhibits making many inferences about lasting consumption effects. Moreover, even if the research were to provide definitive results about behavioral effects, they would have little direct relevance for policy.

Potential effects on investment are a somewhat different story. As in the case of labor supply, there are some general policy preferences. Specifically, if the poor under welfare programs can make investments that lift them out of poverty, that would be desirable. The most likely place for a negative income tax to affect investment behavior is the area of human capital. Human capital investments operate to alter the returns to market work and thus are intertwined with the effects of a negative income tax that also alters the net benefits of market labor. Within the experiments, analyses have considered both the school attendance decisions and the scholastic performance of children in experimental families. The former seems much more relevant for policy purposes.

A negative income tax will lower the costs of continuing schooling, by lessening the cost of not being in the labor force. Further, the reduction in costs observed in the experiment will be the same as that from an ongoing program. The uncertainty in evaluating the experiments and projecting to ongoing programs arises in considering the potential effects on the returns to more schooling. A negative income tax could potentially lower the benefits to more schooling, but this would depend upon the generosity of the program and the potential earnings of the individual with and without any added schooling. For the experimental time period, at least, a negative income tax does appear to induce more schooling. In fact, for youth the reduction in labor supply brought about by the negative income tax is almost perfectly offset by increased school attendance. Thus the encouragement of skill development by youth may be one of the positive sidelights of a negative income tax.

[1]Such a division obviously reflects my economics background. Other taxonomies are plausible, and this taxonomy leaves out a variety of possible concerns such as delinquency rates, political behavior, or psychological factors. See, for example, Rossi (1975), Hannan (1978). While there were some attempts to analyze such noneconomic outcomes, no significant and consistent results emerged. Therefore, the limited focus of this paper does not distort the findings of the experiments.

[2]Incorporating dynamic adjustment processes in models estimated from the experimental data requires imposing an intertemporal structure on the models. In general, given the limited time dimension of the data, this structure cannot be tested or evaluated in any satisfactory manner.

[3]This might be interpreted as donors having preferences over the consumption that results from altruistic transfers.

[4]Almost exclusively, analyses of housing have focused on expenditure relationships, as opposed to real components of housing. This is clearly a result of both measurement difficulties and the heterogeneity of the housing bundle: an increase in the number of bedrooms in housing units is difficult to compare with an improvement in the quality of a unit. It does, however, mean that evaluation is more difficult because it is not possible to ascertain whether increases in housing consumption involve improvements in external conditions (which are most closely related to externality arguments), better space that might be beneficial to the study behavior of children, or other changes.

There is a certain ambivalence in evaluating outcomes on the basis of expenditures. All other things equal, we would surely like the poor to spend less on housing, not more. In fact, the housing allowance experiments evaluated increases in both spending (generally labeled a good thing) and rent burden or proportion of income going to housing (generally labeled a bad thing). In well-functioning markets, we are willing to presume that increased spending connotes improved conditions. However, since the housing allowance at times gives people an incentive to simply spend more even if the quality doesn't change, there are some questions about the interpretation of expenditures.

One way to consider improvements in real quality is through the analysis of hedonic price models. While this was done in both the negative income tax and housing allowance experiments, the findings appear to be quite sensitive to model specification.

[5]Estimates of income elasticities from the experiments tend to be quite low. The precise estimates depend very much on model specification, on the definition of income, and so forth. An elasticity of 0.5 is an estimate related to permanent income of the poor (Hanushek and Quigley 1982). The comparable elasticity from the demand experiments for current income is around 0.2. Comparisons of direct estimates (nonexperimental) and of those from the experiments are found in Hanushek and Quigley 1981.

[6]Ohls and Thomas 1979 do find that income maintenance dollars have a lesser effect on homeownership probabilities than dollars of income from other sources. This may well be a reflection of the discounting of negative income tax payments in individuals' calculations of their permanent income.

[7]The New Jersey experiment provides mixed evidence on homeownership (Wooldridge 1977 and Poirier 1977). In particular, any experimental effects disappear when disaggregated by ethnic group in Poirier's estimates.

[8]As Venti and Wise (1984) point out, the strength of this subsidy depends on whether the person making the schooling decision is a child in a family unit receiving a subsidy or is in a separate household and, in the former case, on the character of household decision-making.

[9]The central studies are: Mallar (1976) for New Jersey; McDonald and Stephenson (1979) for Gary; Weiss, Hall, and Dong (1980); and Venti and Wise (1984) for Seattle-Denver.

[10]Training subsidies were supposedly only for training directly related to occupational or job choices. All discussions of the program operations, however, emphasize that application of this criterion was very loose. Of those accepting subsidies, a majority went to community colleges, but there was considerable variation in this.

[11]Weiss, Hall, and Dong (1980) suggest an effect for heads of household already in school, but a small effect for other youths. Venti and Wise (1984) simply state in a footnote that this subsidy had no effect.

[12]This is an example where the experiments have offered a vehicle for pursuing research that is only tangentially related to the experiments. Because the experiments collected such detailed, longitudinal data on families, they provided key information for

investigating educational performance. By adding a side data collection effort at the schools, a unique data set on schooling was created.

[13]While most people are willing to accept the basic plausibility of these notions, it should be noted that the direct research on these matters does not allow very precise statements about their relationship to scholastic performance.

Further, there is no consistent evidence of health effects or even of increases in health expenditures from the experiments. For example, the findings by Kehrer and Wolin (1979) on birth weights have not been replicated elsewhere.

[14]See Lindert (1977); Belmont and Marolla (1973); Zajonc and Markus (1975); Hanushek (1986a).

[15]Analyses of Seattle-Denver data on school performance—concentrating on home environment—can be found in Manheim and Minchella 1978 and Knickman 1979. Neither finds much in the way of significant home environment effects.

[16]As discussed below, the rural analyses contained much more elaborate measures of potential treatment effects, including among other things interactions of treatments with a variety of measures of family and student characteristics.

[17]The specific output measures analyzed varied by site. For Gary, performance on standardized reading tests, academic grade point average, and days absent were considered; for Iowa and North Carolina, comportment grades were also considered. Absences were significantly reduced and comportment grades were significantly increased in the early grades in North Carolina (Maynard 1977). In Gary, there were some significant differences in academic grade point averages in later grades.

[18]Possible exceptions are analyses of the effects of direct training incentives or, in the case of the housing allowance experiments, estimates of price elasticities of housing demand.

References

Belmont, Lillian, and Francis A. Marolla. "Birth Order, Family Size, and Intelligence," *Science,* 182 (December 14, 1973), pp. 1096-1101.

Bradbury, Katharine L. and Anthony Downs, eds. *Do Housing Allowances Work?* Washington, DC: The Brookings Institution, 1981.

Burtless, Gary, and David Greenberg. "Inferences Concerning Labor Supply Behavior Based on Limited Duration Experiments," *American Economic Review,* 72 (June 1982), pp. 488-497.

Davis, Margaret R., and Kenneth Kehrer. "Overview of Research on Health, Consumption, and Social Behavior," in SRI International. *Final Report of the Seattle-Denver Income Maintenance Experiment; Volume 1: Design and Results,* May 1983, pp. 387-437.

Dynarski, Mark, and Steven M. Sheffrin. "Housing Purchases and Transitory Income: A Study with Panel Data," *Review of Economics and Statistics,* 67 (May 1985), pp. 195-204.

Friedman, Milton. *Capitalism and Freedom.* Chicago: University of Chicago Press, 1962.

Garfinkel, Irwin. "The Effects of Welfare Programs on Experimental Responses," *Journal of Human Resources,* 9 (Fall 1974), pp. 504-529.

Hall, Arden R. "The Counseling and Training Subsidy Treatments," *Journal of Human Resources,* 15 (Fall 1980), pp. 591-610.

Hannan, Michael T. "Noneconomic Outcomes," in John L. Palmer and Joseph A. Pechman, eds., *Welfare in Rural Areas,* Washington, DC: The Brookings Institution, 1978, pp. 183-205.

Hanushek, Eric A. "The Trade-off Between Child Quantity and Quality: Some Empirical Evidence," University of Rochester, 1986. [a]

———. "The Economics of Schooling: Production and Efficiency in the Public Schools," *Journal of Economic Literature,* September 1986. [b]

Hanushek, Eric A., and John M. Quigley. "The Dynamics of the Housing Market: A Stock Adjustment Model of Housing Consumption," *Journal of Urban Economics,* January 1979, pp. 90-111.

———. "Consumption Aspects," in Katharine L. Bradbury and Anthony Downs, eds., *Do Housing Allowances Work?* Washington, DC: The Brookings Institution, 1981, pp. 185-240.

———. "The Determinants of Housing Demand," in J. Vernon Henderson, ed., *Research in Urban Economics,* Greenwich, CT: JAI Press, 1982, pp. 221-242.

Hill, C. Russell, and Frank P. Stafford. "Allocation of Time to Preschool Children and Educational Opportunity," *Journal of Human Resources,* 9 (Summer 1974), pp. 323-341.

———. "Parental Care of Children: Time Diary Estimates of Quantity, Predictability, and Variety," *Journal of Human Resources,* 15 (Spring 1980), pp. 219-239.

Hollister, Robinson G. "Comment," in John L. Palmer and Joseph A. Pechman, eds., *Welfare in Rural Areas,* Washington, DC: The Brookings Institution, 1978, pp. 173-182.

Kaluzny, Richard L. "Changes in the Consumption of Housing Services: The Gary Experiment," *Journal of Human Resources,* 14 (Fall 1979), pp. 496-506.

Keeley, Michael C. "The Effect of a Negative Income Tax on Migration," *Journal of Human Resources,* 15 (Fall 1980), pp. 695-706.

Kehrer, Barbara, and Charles M. Wolin. "Impact of Income Maintenance on Low Birth Weight: Evidence from the Gary Experiment," *Journal of Human Resources,* 14 (Fall 1979), pp. 434-462.

Knickman, James R. "The Influence of Home Environments on Children's School Performance: Evidence from the Seattle-Denver Income Maintenance Experiments," Mathematica Policy Research, November 1979.

Leibowitz, Arleen. "Home Investments in Children," *Journal of Political Economy,* 82 (Mar/Apr 1974), Pt. II, S111-131.

Lindert, Peter H., "Sibling Position and Achievement," *Journal of Human Resources,* 12 (Spring 1977), pp. 198-219.

Mallar, Charles. "Educational and Labor Supply Responses of Young Adults in Experimental Families," in Harold Watts and Albert Rees, eds., *The New Jersey Income Maintenance Experiment. Vol. II: Labor Supply Responses,* New York: Academic Press, 1976.

Manheim, Larry M. and Mary Ellen Minchella. "The Effects of Income Maintenance on the School Performance of Children: Results from the Seattle and Denver

Experiments," Discussion Paper DP-78C-01, Mathematica Policy Research, September 1978.
Maynard, Rebecca A., and David Crawford. "School Performance," in *Rural Income Maintenance Experiment Final Report,* Madison, WI: Institute for Research on Poverty, 1976.
Maynard, Rebecca A., and Richard J. Murnane. "The Effects of a Negative Income Tax on School Performance," *Journal of Human Resources,* 14 (Fall 1979), pp. 463-476.
McDonald, John F., and Stanley P. Stephenson, Jr. "The Effect of Income Maintenance on the School-Enrollment and Labor-Supply Decisions of Teenagers," *Journal of Human Resources,* 14 (Fall 1979), pp. 488-495.
Metcalf, Charles E. "Predicting the Effects of Permanent Programs from a Limited Duration Experiment," *Journal of Human Resources,* 9 (Fall 1974), pp. 530-555.
________. "Making Inferences from Controlled Income Maintenance Experiments," *American Economic Review,* 63 (June 1973), pp. 478-83.
Michael, Robert T. "The Consumption Studies," in John L. Palmer and Joseph A. Pechman, eds., *Welfare in Rural Areas,* Washington, DC: The Brookings Institution, 1978, pp. 149-171.
Moeller, John F. "Consumer Expenditure Responses to Income Redistribution Programs," *Review of Economics and Statistics,* 58 (August 1981), pp. 409-421.
Murnane, Richard J., Rebecca A. Maynard, and James C. Ohls. "Home Resources and Children's Achievement," *Review of Economics and Statistics,* 58 (August 1981), pp. 369-377.
O'Connor, Frank J., and J. Patrick Madden. "The Negative Income Tax and the Quality of Dietary Intake," *Journal of Human Resources,* 14 (Fall 1979), pp. 507-517.
Ohls, James C., and Cynthia Thomas. "The Effects of the Seattle and Denver Income Maintenance Experiments on Housing Consumption, Ownership, and Mobility," Draft Report, Denver: Mathematica Policy Research, January 1979.
Poirier, Dale J. "The Determinants of Home Buying," In Harold W. Watts and Albert Rees, eds., *The New Jersey Income-Maintenance Experiment. Vol. III: Expenditures, Health, and Social Behavior; and the Quality of the Evidence,* New York: Academic Press, 1977, pp. 73-92.
Rea, Samuel A., Jr. "Investment in Human Capital Under a Negative Income Tax," *Canadian Journal of Economics,* 10 (November 1977), pp. 607-620.
Rossi, Peter H. "A Critical Review of the Analysis of Nonlabor Force Responses," in Joseph A. Pechman and P. Michael Timpane, eds., *Work Incentives and Income Guarantees: The New Jersey Negative Income Tax Experiment,* Washington, DC: The Brookings Institution, 1975, pp. 157-182.
SRI International. *Final Report of the Seattle-Denver Income Maintenance Experiment; Volume 1: Design and Results,* May 1983.
Venti, Steven F., and David A. Wise. "Income Maintenance and the School and Work Decisions of Youth," mimeo, Harvard University, March 1984.
Weiss, Yoram, Arden Hall, and Fred Dong. "The Effect of Price and Income on Investment in Schooling," *Journal of Human Resources,* 15 (Fall 1980), pp. 611-640.
West, Richard W. "Effects on Wage Rates: An Interim Report," *Journal of Human Resources,* 15 (Fall 1980), pp. 641-653.
Wooldridge, Judith. "Housing Consumption," in Harold W. Watts and Albert Rees, eds., *The New Jersey Income-Maintenance Experiment. Vol. III: Expenditures, Health, and Social Behavior; and the Quality of the Evidence,* New York: Academic Press, 1977, pp. 45-72.
Zajonc, R.B., and Gregory B. Markus. "Birth Order and Intellectual Development," *Psychological Review,* 82 (January 1975), pp. 74-88.

Discussion

*Katharine L. Bradbury**

Research from the income maintenance experiments relating to subjects other than labor supply and family composition seems to suffer from an inferiority complex. One sees this in similar summary papers from previous conferences as well as in Eric Hanushek's paper for this conference. Although the design and research of the income maintenance experiments focused on labor supply and much research also looked at the family composition effects of income maintenance, what Hanushek calls the "residual" analyses add to our knowledge in many ways. In particular, they may have important policy implications, from informing concerned taxpayers how income maintenance benefits are spent, to helping to choose between in-kind and cash benefits. The relevance of these findings is heightened in the current policy context where both the right and the left have focused attention on the "culture of poverty" as a major stumbling block to efforts aimed at improving the opportunities and well-being of the poor.

Hanushek presents the experimental findings regarding housing and education. These comments will briefly summarize the research relating to some other areas of consumption and investment, including health, as well as the social and psychological investigations.

Why should we be interested in the effects of income maintenance on behavior other than labor supply and family composition? William Baumol laid out the issue very clearly in an early paper summarizing the consumption, health, and social behavior results of the New Jersey-Pennsylvania experiments. He said:

> Those who fear the worst of a [negative income tax] system may hold the hypothesis that a large part of the payments will be wasted by the recipients—

*Economist, Federal Reserve Bank of Boston.

> either being spent on drugs, drinks, and gambling or being dissipated in increased leisure time unproductively used. Those espousing such an extreme view would say that a program of unconditional cash payments should be avoided, perhaps to protect the potential recipients from their own folly, and certainly to prevent the use of public funds for such purposes.
>
> At the other extreme is the view that such a program might well improve the life style of the recipients. The assurance of financial support and the increase in expected income might lead to a modification in attitudes that could, perhaps, be described as the adoption of middle-class values and a reconciliation with the goals of the bulk of society. Such attitude changes might be expected to lead, among other things, to political activity, increased interest in education and quality of neighborhoods, lower crime rates, and reduced neurosis and psychosis.

While economists prefer to avoid making value judgments about how people spend their money, taxpayers appear to have no such reluctance. The experimental evidence on consumption provides some facts that can replace the stereotypes which seem to dominate in taxpayer reactions.

A second reason to be interested in these results is that they may provide useful information to policymakers faced with a choice between cash assistance and in-kind or targeted forms of aid. For example, the finding of lower income elasticities of rental expenditures in the income maintenance experiments than in the housing allowance experiments led Seattle-Denver researchers to conclude that targeting has important benefits if the aim is to encourage housing consumption.

Aside from labor supply and family composition, the experimental results are fairly easily summarized: recipients of cash payments do not change their behavior noticeably except in reflection of their increased income. Some specific results are described below.

Regarding consumption, increased income, in general, enabled recipients of negative income tax payments to consume more and they did so in rough proportion to their consumption in the absence of such payments. In other words, the source of income is not particularly important in its allocation among possible uses, and income levels were not increased so much by the experiments that basic consumption patterns shifted in response. The results are mixed in terms of whether payments were used to increase net assets and net savings other than through changes in housing, although net worth certainly rose for some subgroups in some sites. The limited duration of the experiments raises the question of permanent versus transitory income changes and hence issues of timing with respect to the accumulation of durables, which the researchers investigated. But in any case, the experiments do not suggest that payments were used in ways that taxpayers would view as purely frivolous or immoral.

The area of health is of particular interest to those concerned with welfare policy, since poor health is thought to be one of the immediate causes of poverty and one of the ways poverty is perpetuated from generation to generation. It is difficult to measure actual health status, so most of the analyses focused on such measurables as utilization of medical care, nutritional adequacy, and infant birth weight.

As far as the researchers could determine, medical care utilization did not increase and health status did not improve as a result of income maintenance payments. A major caveat regarding the finding of no effect is that in most cases investigators were unable to control for the availability of Medicare or health insurance, which could confound the experimental effect. In the Rural Experiment, children's health status appeared to improve, but the differences were not statistically significant. Some evidence surfaced in New Jersey that recipients altered patterns of medical care utilization, shifting toward private physicians from hospitals and clinics. In Seattle and Denver, recipients spent slightly more on health care than the control group. A general conclusion one might draw from these results is that programs aimed directly at health care have a better chance to have an effect on health status than do cash transfers.

Low infant birth weight can be an indicator of poor health in the mother, and is often associated with later poor health and developmental difficulties of the child. Birth weight was studied only in the Gary and Seattle-Denver experiments. Seattle-Denver researchers found no significant effects of experimental status on low birth weight and hence infant health status. In contrast, payments in Gary were associated with significant declines in the prevalence of low birth weight among those mothers at the highest risk. The researchers found no difference in the frequency or type of prenatal care received by experimental and control mothers in Gary, so they hypothesized that the improvement resulted from improved nutrition, although they had no direct evidence to test the hypothesis.

Nutrition was directly studied only in the Rural Experiment. There was no ascertainable effect of payments on nutrition in Iowa, but a small persistent positive effect in rural North Carolina, where the baseline nutrition levels were noticeably more deficient. This positive effect showed up for 9 of 10 basic nutrients examined. The researchers inferred that those families with deficient nutrition used their payments to bring nutritional levels closer to minimally adequate levels, while those with adequate nutrition pursued other goals with their increased income.

Researchers also studied a number of measures related to recipients' attitudes, mental health, community involvement, political activity, social integration, and the like. In the Rural Experiment, researchers concluded that payments had no negative effects on psychological well-being and overall perhaps some slightly positive effects. In addition, experimentals were more likely to vote and otherwise participate in electoral activities than controls. In New Jersey, researchers found no effects on psychological distress or social integration. While they failed to find that reduced levels of distress were associated with additional income, they point out that they also failed to find any psychologically deleterious effects on participants, such as a decline in self-esteem. In the Seattle-Denver studies, a few subgroups of experimentals showed slightly higher levels of psychological distress than controls, but the conclusion again was "no effect."

Overall, these results suggest that the lives of recipients were not dramatically altered by the payments offered for a limited time period in the income maintenance experiments. Consumption rose modestly, as would be expected with a modest rise in income. Most other indicators of well-being showed little if any change. In terms of long-term consequences, the improvements in education (noted in the Hanushek paper) and the improvements in birth weight and nutrition among those with the worst initial deficits are probably the most promising. In these areas there is no normative confusion about what constitutes an improvement.

These results allow us to reject, as Baumol did, his two extreme hypotheses: that payments will be "squandered" or that recipients will be transformed into members of the middle class. Since these views are part of the underlying differences between camps in the welfare reform debate, the lack of noticeable effects is, in itself, a notable finding.

References

Baumol, William J. ''An Overview of the Results on Consumption, Health, and Social Behavior,'' *Journal of Human Resources,* 9 (Spring 1974), pp. 253-64. Quote from p. 253.

Davis, Margaret R., and Kenneth C. Kehrer, ''Overview of Research on Health, Consumption, and Social Behavior,'' *Final Report of the Seattle/Denver Income Maintenance Experiment, Volume 1,* Part VI, pp. 385-446.

Hannan, Michael T. ''Noneconomic Outcomes,'' in John L. Palmer and Joseph A. Pechman, eds., *Welfare in Rural Areas: The North Carolina-Iowa Income Maintenance Experiment,* Washington, DC: The Brookings Institution, 1978, with comments by Stanley H. Masters and Robinson G. Hollister.

Kehrer, Barbara H., and Charles M. Wolin. ''Impact of Income Maintenance on Low Birth Weight: Evidence from the Gary Experiment,'' *Journal of Human Resources,* 14 (Fall 1979), pp. 434-62

Kerachsky, Stuart H. ''Health and Medical Care Utilization; a Second Approach,'' in Harold W. Watts and Albert Rees eds., *The New Jersey Income Maintenance Experiments, Volume III,* Academic Press, 1977.

Ladinsky, Jack, and Anna Wells. ''Social Integration, Leisure Activity, Media Exposure, and Lifestyle Enhancement,'' in Watts and Rees, *op. cit.*

Lefcowitz, Myron J., and David Elesh. ''Health and Medical Care Utilization,'' in Watts and Rees, *op. cit.*

Metcalf, Charles E. ''Consumption Behavior: Implications for a Permanent Program,'' in Watts and Rees, *op. cit.*

Michael, Robert T. ''The Consumption Studies,'' in Palmer and Pechman, *op. cit.,* with comments by Robert J. Lampman and Aage B. Sorensen.

Middleton, Russell, and Vernon L. Allen. ''Social Psychological Effects,'' in Watts and Rees, *op. cit.*

Nicholson, Walter. ''Expenditure Patterns: a Descriptive Survey,'' in Watts and Rees, *op. cit.*

O'Connor, J. Frank, and J. Patrick Madden. ''Communication on the Rural Experiment: The Negative Income Tax and the Quality of Dietary Intake,'' *Journal of Human Resources,* 14 (Fall 1979), pp. 507-17.

Discussion

*Robert T. Michael**

Hanushek has given us a thoughtful essay about the studies from the negative income tax experiments on subjects other than labor supply and marital behavior. His decision, however, to concentrate on two topics of interest to policymakers leaves many of the studies by the research teams unreferenced here. Consequently, Katharine Bradbury and I, as discussants, independently chose to amplify Hanushek's comments rather than critique them. Fortunately, we selected different subsets of results for discussion and so among the three of us we may have provided a reasonably complete sketch of the studies undertaken.

I wish to make two general points in these remarks. First, important suggestive results in a variety of areas should not be ignored in any overall review of what was learned from the negative income tax experiments. Second, a negative income tax experiment, by its nature, is ill-suited to yield high quality data for the analysis of a wide range of behavior. The remainder of my remarks will elaborate on these two points.

Hanushek summarizes the results on housing and schooling under the appropriate taxonomy of consumption and investment. He indicates that his selection of results was made in part on the basis of "common findings that might be generalized." Additional findings in one or another of the four experiments, not reinforced by the other experiments, should also be mentioned. The absence of corroboration of findings from other experiments was often—not always—attributable to a difference in design, circumstance or method of analysis.

On the topic of health, which Katharine Bradbury has reviewed, I

*Director, National Opinion Research Center and Professor of Education, University of Chicago.

wish only to emphasize an important suggestive finding from the North Carolina rural experiment. Among the very poor in that survey the experiment seemed to have small but consistently positive effects on the nutrient intake of the family members, which appeared to be stronger over time.[1]

Child care was studied in the Seattle-Denver income maintenance experiments (see Kurz, Robins, and Spiegelman 1975, and Robins and Spiegelman 1978). Substitution toward market forms of child care was observed, replacing family care and other forms of nonmarket care.

Migration is another subject investigated using the Seattle-Denver data. That experiment, unlike its predecessors, permitted the families that received payments in the Seattle or Denver area to move and retain their rights to receive the payments. Keeley (1980) found the rate of migration nearly 50 percent higher for those in the experiment than for the controls. Keeley found as well that the locations to which these families moved were relatively rich in amenities, such as low variation in temperature or relatively high average January temperature. As he suggests, this is what one might expect, given the negative income tax's high rate of taxation on money income relative to amenities.

Keeley (1980) investigated the fertility effects of the negative income tax using the Seattle-Denver data. For couples married at the outset of the experiment, the effects differed by race-ethnic group and duration of the experiment. For whites the effect was negative for the group in the five-year duration experiment, but for Hispanics it was positive for the three-year duration group. There was no statistically significant effect for blacks. Keeley points out that the net effect of a negative income tax on fertility is not clear in theory, because of potentially competing income, price and subsidy effects. The lack of consistency across these three groups of married women leaves us with no convenient generalization about fertility effects, but Keeley's paper suggests the effects may be more complex than his model was able to sort out. For women not married at the outset, Keeley reports no discernible effect of the experiment on fertility.

In all, the income maintenance experiments had a wide range of results on consumption of specific items, on investment in children, migration, health and other forms of human capital, and on other aspects of social behavior and attitudes. These results are not in general strong or clearcut.[2] Elsewhere (Michael 1978) I have criticized the studies from the rural experiment for focusing far too single-mindedly on the regression coefficient on the experimental effect. In many cases the opportunity to provide some useful descriptive information about the qualitative nature of the lives of these low-income families was lost in the rush to report a negative income tax coefficient. The research results on consumption and other non-labor-supply, non-marital effects from these experiments are not one of the best features of the experiments. The results are broader than Hanushek's paper would suggest, but they are generally weak methodologically. They mostly report mixed, often puzzling and inconsistent findings, and they do not offer us a good guide to the various demographic, social and economic (non-labor-supply) repercussions of introducing a nationwide negative income tax.

One of the reasons for the weak results in these studies is implied by my second point: an income maintenance experiment is poorly suited for the study of a wide spectrum of social behavior. This is true for at least two reasons. First, the conduct of the experiment is sufficiently taxing that a secondary purpose, such as the collection of high-quality data on consumption or social behavior, is not given adequate resources.

Income maintenance experiments are by their nature quite costly. In addition to the cost of planning, conducting and analyzing the survey, there is the cost of the transfer income. Also, the funding agency interacts frequently with the survey and analysis teams in a social experiment, and that interaction tends to keep attention focused on the policy-relevant issue. As the explicit purpose of the experiment is to observe labor supply response, that response deservedly receives most of the attention of the planners. Any effort to divert resources of time or money to ancillary topics is resisted.

Another factor is political involvement, both with Congress and with the local agencies of government whose cooperation in the experimental survey enterprise is so essential. Given the costliness, the funding-agency involvement and the political oversight, it is little wonder that the inherent riskiness and high stakes of a negative income tax experiment keep the survey organization in a state of some anxiety throughout the experiment. That tenseness does not foster a receptive atmosphere for suggestions about improving the quality of data on secondary research topics.

On top of these pressures is the second obstacle to the study of these ancillary topics: the sampling for an income maintenance experiment requires the selection of a control group and an experimental group and necessarily these are concentrated in one or a few localities. Neither of these features is useful from the point of view of most secondary study topics.

Other large-scale studies using omnibus questionnaires and longitudinal surveys have proven themselves excellent vehicles for adding on topical modules of high quality. The pressures on the experimental surveys, however, coupled with their sampling design, do not, I am convinced, foster good data on these less essential topics.

If my second point is correct, the diversity of research on non-labor-supply effects in the negative income tax experiments is quite remarkable and reflects well on the ingenuity and intellectual curiosity of scholars involved in their design and analysis. Yet, the generally weak and often conflicting results in these areas reflect, I am afraid, the quality of the data on these subjects. One does not see the experimental data on these other topics used subsequently in general analyses, a further indication of their quality. The data tapes are not typically placed in national data archives after the project report is prepared, and the general research community has shown little interest in obtaining these data for general analyses. Any assessment of the negative income tax experiments should include the fact that the data typically are not of very great value for studies other than their principal, explicit purpose.

---

[1] See O'Connor, Madden and Prindle 1976. In an essay summarizing the consumption studies from the rural experiment, I suggest reasons why the results in that study may understate the impact of the income transfer on the health-relevant nutritional intake. See Michael 1978.

[2] For a good review of the Seattle-Denver income maintenance experiments results see Davis and Kehrer 1983; the Fall 1980 issue of *The Journal of Human Resources* has several summary papers from the Seattle-Denver experiments as well.

References

Davis, Margaret R., and Kenneth C. Kehrer. "Overview of Research on Health, Consumption, and Social Behavior," in *Final Report of the Seattle-Denver Income Maintenance Experiment, Vol. 1, Design and Results,* SRI International, U.S. Government Printing Office: May 1983, pp. 385-466.

Keeley, Michael C. "The Effect of a Negative Income Tax on Migration," *The Journal of Human Resources,* 15 (Fall 1980), pp. 675-706.

Kurz, M., Philip K. Robins, and Robert G. Spiegelman. "A Study of the Demand for Child Care of Working Mothers," SRI Research Memorandum #27, August 1975.

Michael, Robert T. "The Consumption Studies," in John L. Palmer and Joseph A. Pechman, eds., *Welfare in Rural Areas: The North Carolina-Iowa Income Maintenance Experiment,* The Brookings Institution, 1978, pp. 149-171.

O'Connor, J. Frank, J. Patrick Madden, and Allen M. Prindle. "Nutrition," in D. Lee Bawden and William S. Harrar, eds., *Rural Income Maintenance Experiment: Final Report,* University of Wisconsin, Institute for Research on Poverty, 1976.

Robins, Philip K., and Robert G. Spiegelman. "An Econometric Model of the Demand for Child Care," *Economic Inquiry,* 16 (January 1978), pp. 83-94.

Evaluating the Methodology of Social Experiments

*Arnold Zellner and Peter E. Rossi**

In view of the many papers and books that have been written analyzing the methodology of the income maintenance experiments as well as other social experiments, it is indeed difficult to say anything entirely new. However, we shall emphasize what we consider to be important, basic points in the planning, execution and evaluation of social experiments that may prove useful in future experiments. The plan of the paper is as follows. In the next section, we put forward considerations relevant for evaluating social experiments. We then take up design issues within the context of static designs, while the following section is devoted to issues that arise in dynamic contexts, the usual setting for most social experiments. Suggestions for linking social experiments to already existing longitudinal data bases are presented and discussed. In both static and dynamic contexts, we discuss the roles of models, whether statistical or structural, and of randomization. Design for prediction of relevant experimental outcomes is emphasized and illustrated in terms of simplified versions of the negative income tax experiments. Finally, we present a summary and some concluding remarks.

Considerations Relevant for Evaluating the Methodology of Social Experiments

Since social experiments usually involve the expenditure of millions of dollars, resources that have alternative research uses and potentially great social value, it is critical that the experiments be conducted in a

---

*Professor of Economics and Statistics and Assistant Professor of Econometrics and Statistics, respectively, Graduate School of Business, University of Chicago. Michael A. Zellner provided helpful research assistance.

manner that is methodologically sound. The task of defining good or optimal methodology is difficult, however, because social experiments are multidimensional in nature, involving many disciplines and activities. Also, good experimentation involves creativity and innovation, which are difficult to define. We will discuss critical features that can be of vital importance for the success of social experiments.

A clear-cut statement of the objectives of an experiment is the first requirement of a good methodological approach. Poorly formulated objectives are an indication of poor methodology in general. If an experiment is purely for research, then the researchers have the responsibility for formulating its objectives. On the other hand, if the experiment has mainly policy objectives, then it is critical that researchers and relevant policymakers jointly formulate the objectives of the experiment.[1]

Once the objectives of an experiment have been clearly formulated, the second step involves a feasibility study, in order to determine how and if the objectives can be realized. This should include a review of previous studies and data, experimental and nonexperimental, relating to the objectives of the current experiment. It should also consider in detail the needed inputs for the proposed experiment. Usually subject matter specialists, well versed in the subject to be investigated,[2] survey experts, and design statisticians will be required. Most important is the development of an operational approach that is capable of realizing the objectives of the experiment.

If the objectives involve the production of results in a short time, the feasibility study may indicate that calculations using nonexperimental data are all that can be done. On the other hand, if a social experiment seems feasible, its design and costs should be explicitly developed. Finally, the quality of both the research team and the managerial or administrative personnel is of key importance.

In the feasibility study, it is desirable that calculations be performed to provide preliminary estimates of important effects.[3] These rough calculations provide important order-of-magnitude estimates that can be quite useful as background information in evaluating experimental designs. Last, it is usually good practice to execute a "pilot" or "test" trial of the experiment, just as survey questionnaires are subject to pretests. Such pilot experiments can reveal many unexpected results and aid in the redesign of the "final" experiment.

The quality of measurements is a third key issue in the evaluation of the methodology of social experiments.[4] If measurements are of low quality, the results of an experiment are of dubious value. Are all appropriate and relevant variables being measured? Are the measurements afflicted by response and recall biases? Do subjects misrepresent data for various reasons? Are Hawthorne effects present? Checks on the validity of the basic data provided by an experiment must be pursued vigorously in a good methodological approach to social experimentation. This requires that data specialists and those familiar with measurement methodology be involved in the execution of a social experiment.

Fourth, as stated above, outstanding subject matter specialists are required, in order to ensure that the methodology of an experiment is appropriate.[5] An experiment usually involves subjecting experimental units to important changed conditions. Since their responses to the changed conditions are usually adaptive and dynamic in nature, care must be taken in choosing a model that can represent such responses and serve as a basis for choice of an experimental design. Designs can be chosen not only to estimate effects from given models but also to provide data that are useful for testing uncertain features of a model brought to light by experts' analyses of existing theory and models. For example, such considerations may involve use of a model with several equations rather than a single equation for, say, labor supply. If the multiple equation model is appropriate, a design based on a single equation is inappropriate and can lead to erroneous conclusions.[6]

Fifth, the design of the experiments and other statistical issues are basic to a good methodological approach. If the objectives of an experiment involve generalization of the results to an entire population, then the sample of experimental units has to be a sample from the relevant population.[7] The relevant population must be carefully defined with respect to spatial, temporal[8] and other characteristics. Further complications arise from the inability of experimenters to require participation in an experiment. Those volunteering to participate may possibly be different from those not willing to participate and if so, the experiment may be subject to selection biases. (See Duan et al. (1984) for an evaluation of models that attempt to correct for selection bias.) Further, there is the problem of attrition.[9] It is important to do everything possible to keep the attrition rate low. Sampling the dropouts and using the results of analyses relating to these samples is one way of checking on the importance of attrition bias and of correcting for it. Constant vigilance with respect to possible sources of bias and the use of every means possible to avoid such biases are characteristic of a good methodological approach.

Assigning units to treatment and control groups at random is considered good practice by most experimenters. However, good randomization procedures depend on an intimate knowledge of the model generating the observations, as Jeffreys (1967, p. 239ff.) and Conlisk (1985) have demonstrated. For example, Conlisk (1985) has shown that effective randomization when treatment effects are additive is not effective when treatment effects enter a model nonlinearly. Thus, how one randomizes depends on what one knows about features of a model for the observations—for example, see Rubin (1974). Also, a randomized design for a static model may not be effective if the static model misrepresents dynamic responses to ''treatments'' and other variables. Last, it is worthwhile to emphasize not only precision in design but also balance and robustness of design, as Morris (1979) has emphasized.

Sixth, a successful social experiment requires good managers and administrative methods. A good research manager will be invaluable in scheduling operations of a large-scale social experiment, keeping control of costs, instituting good data management procedures, and, most importantly, guiding the project so as to raise the probability of its success in meeting its objectives. Good management involves not only selection of appropriate researchers and other personnel but also surveillance of the project to ensure that researchers are pursuing the stated objectives of the experiment. There usually is a great danger that researchers may get involved in tangential problems and issues and possibly provide the right answer to the wrong problem, a statistical error of the third kind.[10]

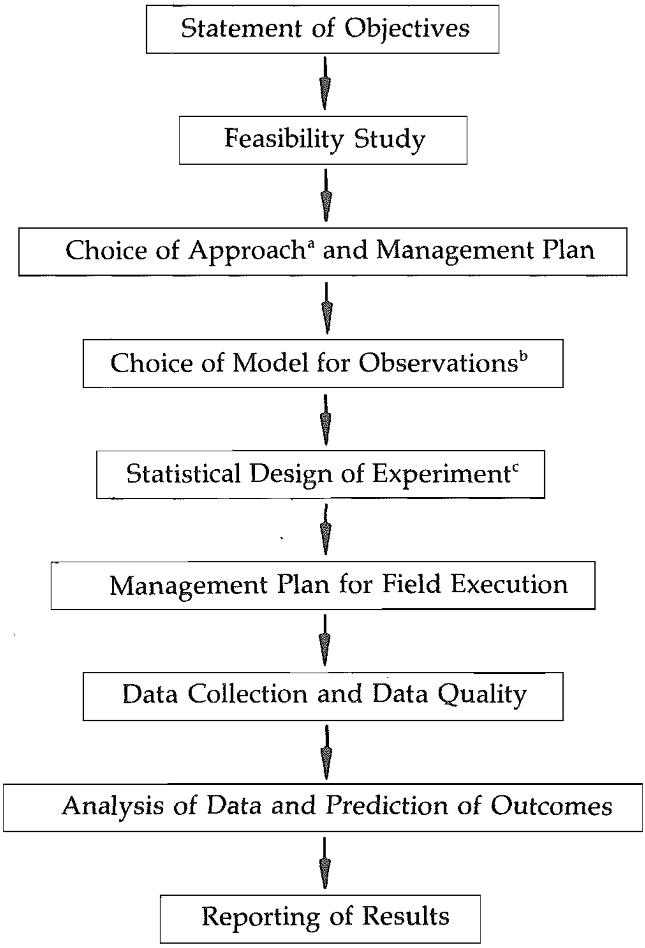
As the experience with the negative income tax experiments has indicated, data collection and data processing costs have been a large fraction of total costs of social experiments.[11] Thus, it is important to put a great deal of emphasis on the design of efficient computerized data management systems and on ways to record basic data directly into computerized data bases.

Seventh, good experimental research methodology involves concentration on the prediction of observable outcomes of an experiment and on establishing reproducible results. Predicting the observable outcomes of a social experiment, for example the costs of variants of a negative income tax program, might be the main objective of a social experiment. If so, the experiment should be designed to provide accurate enough predictions of the costs of these variants along with measures of predictive precision. Calculations yielding these predictions should be reproducible. However, the final proof will be the extent to which the experiment's predictions agree with the actual cost of a negative income tax program, if instituted. That is, good experimental methodology results in predictions that are verified in practice.

Finally, the results of an experiment must be well reported. Careless reporting of experimental results can lead to the adoption of incorrect policies, particularly when the experiment's objectives are not one-to-one with policymakers' objectives.[12] Good methodology requires great care in reporting the results of experiments with attention paid to their uncertainties and measures of precision. Results must be reported in a probabilistic framework that policymakers can understand.

The schematic diagram may be useful in providing an overall appreciation of the process of social experimentation, although it should be recognized that possibly important feedback loops have not been included. For example, at various stages in the execution of an experiment research findings may indicate a need for changes in objectives, the model chosen for the observations, or other aspects of the experiment. While some of these eventualities may be taken into account in sequential designs, taking appropriate measures to deal with the unforeseen is an important element of a good methodological approach. (Watts and Bawden (1978) provide an account of some surprises that actually occurred in the experiments.)

Elements Involved in Planning and Executing Social Experiments



---

[a]It is assumed that a social experiment approach is feasible and selected.

[b]It is assumed that a relevant population has been defined.

[c]The statistical design can involve checks on the adequacy of alternative models for the observations and reflect what is learned from "pilot" or "test" trials of the experiment.

Design and Analysis in a Static Framework

A substantial literature has developed on the econometric problems associated with the design of social experiments and the analysis of experimental data. We shall not attempt an exhaustive summary or critique of the econometric literature on social experiments.[13] Rather, in this section, we shall discuss some key features of design of social experiments in the context of a simple "static" setting. The social experiment will take place during one experimental "period," which is assumed to be long enough for households to respond fully to the experimental treatment. The simple static setting will allow us to stress statistical problems common to most social experiments without detailed exposition of complicated structural models. Of course, use of a static model abstracts from some important dynamic considerations in both the design and analysis of experiments, which we will take up in a later section.

Review and Critique of Existing Design Procedures

The design of social experiments involves three distinct sorts of decisions: 1) the choice of experimental population, 2) the choice of the design space or range of possible treatments, and 3) the allocation of subjects (typically, households) to various treatment and control groups. In all major social experiments, specific sites (most often SMSAs) have been chosen for an experiment. On each site, the eligible population is determined and a sample from this population is invited to participate in the experiment. High administrative and field interview costs are cited as reasons for using a site-based approach.[14] The goal of a representative site (Middletown, USA) has proved elusive. As a result, the experimental population is frequently not representative of the target population for a national program. In order to extrapolate the findings of the experiment to a national scale, some sort of structural model must be assumed.[15]

National probability samples should be considered for future experiments. The problems with administration and field operations of a national experiment appear to have been exaggerated by the planners of the negative income tax experiments. A number of national survey organizations have routinely conducted national surveys since the 1960s. (The National Opinion Research Center at the University of Chicago and the Institute for Social Research at the University of Michigan are examples.) Differences in local welfare laws make it all the more necessary to diversify the sample to include more than a handful of sites.[16] In the next section, we propose the formation of an ongoing panel based on a national sample frame for use in social experiments. It should be noted that marketing researchers have used both the national sample approach (Nielsen surveys) and the site-based approach (most often used in the test-marketing of products) to evaluate the effects of changes in advertising and the response of consumers to new products.

Once a target population is selected, the designers of the experiment must select the range of possible treatments. In the negative income tax experiments, the possible combinations of tax rates and support levels determine the design space. The possible range of treatment variables appears to have been selected in an ad hoc fashion through a combination of political compromise and personal judgment. Describing the New Jersey experiment, Conlisk and Watts write:

> The problem, then, was one of specifying a sample in the three dimensional design space of (g,t,w) triplets [support, tax rate, and wage rate]. Sampling was restricted to a region within the design space which provided substantial variation . . ., but kept to (g,t,w) combinations of actual policy interest . . . So the design problem reduced to finding optimal numbers of families . . . to allocate to each design point.[17]

Keeley and Robins (1980b, p. 328) point out in their excellent critique of the design of the Seattle-Denver income maintenance experiments that it is not clear how to "specify the design space" and, more importantly, that "efficiency may be increased by changing the design space for a given response function." The problem of choice of design space is not unique to either the Conlisk/Watts response surface approach (used in all negative income tax experiments) or the ANOVA model approach (used in the design of the Housing Allowance Demand Experiment) to experimental design. Neither approach gives adequate guidance in the choice of the range of treatments.

The designs used in the negative income tax experiments tended to be overly conservative with less variation in treatment than is desirable, particularly when the goal of the experiment is to estimate quantities that are imprecisely known to start with. For example, in the Seattle-Denver experiment, marginal tax rates of between 0.5 and 0.8 were used. The range of these rates is very limited and does not include either the high or the low tax rates that might be expected to produce the most or least labor supply reduction. It is precisely from such extreme experimental conditions that most can be learned about model parameters and model adequacy. In the Seattle-Denver experiments, a tax rate of 0.3 and a support level of $3800 would have had a grant break-even level of $12,667, only slightly higher than the highest break-even level in the study ($12,000).

The distribution of treatment points over the feasible set of treatments is critical for ensuring appropriate precision in the estimation of treatment effects and for testing the model specification. Without an adequate range of treatment levels, precise treatment effect estimates can only be obtained from huge samples, beyond the means of social experimenters. Perhaps most importantly, without a fairly uniform distribution of treatments across the design space it is difficult to test for model misspecification by comparing alternative models. We find little attention directed toward this important problem in the social experimentation literature.

Allocation of participating families across treatment groups has received considerable attention. With few exceptions, economists have rejected a classical analysis of variance approach in favor of a model-based approach developed by Conlisk and Watts. (See Conlisk and Watts (1979) for a full description of this technique.) A response surface-based allocation of households to treatment groups was used, usually with modifications, in all the negative income tax experiments, the health insurance experiment, and several time-of-day electricity pricing experiments. The heart of the response-surface approach to design is a demand equation: in the case of the negative income tax experiment, a demand for leisure equation is used. Given a specification of the demand equation, optimal allocations of households across treatment groups are derived. The goal of the experimental design procedure is assumed to be the maximization of the precision of estimation of the coefficients of the response model, subject to a budget constraint. Optimal allocations from the demand equation model provide non-orthogonal experimental designs, which should provide greater precision in estimation of key model parameters than traditional orthogonal designs as well as samples with lower experimental cost.

A critical assumption in the application of the response surface models is that experimental observations may be more costly than control observations both in terms of benefits and administrative costs. In their excellent critique of the Conlisk/Watts model, Keeley et al. (1980, p. 328) indicate that in the Seattle-Denver experiments experimental observations were assumed to be four times as expensive as controls, when the actual cost ratio was 2.3. Because of the asymmetry of observation costs, the response surface approach yields non-orthogonal and non-randomized experimental designs. Keeley and Robins (1980a, b) and Hausman and Wise (1985) have emphasized the severe problems produced by the endogenous stratification induced by this cost function. It is difficult to understand this preoccupation with cost-effective designs when most of the significant costs of negative income tax experiments are fixed or at least constant across households. Typically, the data processing, field operations and analysis budgets for negative income tax experiments far exceeded the total benefits paid out. (See, for example, the cost data in footnote 11.)

Perhaps the most serious defect of the response model approach is the extreme sensitivity of the optimal designs to model misspecification. As Conlisk (1973) and Aigner (1979a) have reported, optimal designs for one response function can be very suboptimal for other response functions. Given the considerable uncertainty about the appropriate specification of labor supply behavior, a good design procedure should incorporate some robustness to departures from model assumptions. Model misspecification can include incorrect functional form for the demand equation, measurement errors in the independent variables, omitted or latent variables, incorrect distributional assumptions (outliers, sample truncation and censoring), sample selection bias, and incorrect dynamic specifications. Due to the inherent difficulties in measuring income and wage rates and the field operations mistakes made in the negative income tax experiments, the problems of accounting for measurement error biases are particularly important. Measurement error problems are further compounded when the design is stratified based on an endogenous variable that is measured with error. Explicit consideration of measurement errors is a necessary condition for optimal design in these situations.

A natural way to produce robust statistical designs is to use randomization procedures, by which households are randomly assigned to treatment groups according to simple or stratified random sampling procedures. Randomization procedures have been widely used in experimental design in medical, psychological and educational research for many decades. Based on both practical experience and extensive theoretical research (Fisher, 1925 and Kempthorne, 1952) randomization procedures have been shown to have great value in reducing bias in the determination of experimental effects when response models are misspecified. As Hausman and Wise (1985) and Morris (1979) have pointed out, to the extent that unobserved variables are correlated with observed variables over which the design is randomized, the effects of model misspecification are mitigated.[18]

The severe problems that have plagued the response model approach to design have prompted some economists to advocate designs based on the simplest sort of analysis of variance (ANOVA) model. (See, for example, Hausman and Wise, 1985.) Optimal design in an ANOVA framework requires an orthogonal layout of treatments and a random assignment of participants to the various treatment groups (each group corresponds to a row of the X matrix in the Conlisk/Watts model). Of course, the ANOVA model is a special case of the general linear response model used by Conlisk and Watts. It appears that the chief benefits of the ANOVA approach are simplicity of analysis and randomization over participant characteristics, which avoids the problem of endogenous stratification. However, the ANOVA approach is sensitive to model misspecification, as pointed out by Conlisk (1985), requires many observations when there are large numbers of cells, and, most importantly, cannot be used to generalize in important ways from the experimental experience. That is, ANOVA models are designed to test for experimental effects and cannot be easily adapted for predictive purposes. Also, given the problems of a site-based sample and participation and attrition biases, it is highly unlikely that experimental data will be representative of the national target population for a social program. Thus, a response model would have to be built to extrapolate the ANOVA experimental findings to a national scale.

An essential problem with both response function and ANOVA approaches to experimental design is that research objectives of a study are not explicitly included in the objective function used to determine the optimal design. If the objective of a negative income tax experiment is to estimate accurately the cost of a national program, the objective function should be formulated to measure the precision of cost estimates. Similarly, if the goal is to refine estimates of substitution effects, the precision of these estimates should define the objective function.[19] The usefulness of the current experimental design techniques can be gauged by noting that we know of no social experiment in which the original design model was used in the analysis of experimental data. In all of the negative income tax experiments the original response model was discarded in favor of more restrictive labor supply functions, which have drastically fewer parameters.

The unusual role of controls in social experiments also involves difficulties with current design techniques. A control subject is defined in the classical experimentation literature as a subject who received no treatment. In the negative income tax experiments, many control families received current AFDC benefits (of course, the investigators could not ask these households to give up income support altogether) while other families received no welfare benefits of any kind. These control AFDC families are receiving a different treatment, not the null treatment. In fact, the control households are treated in most analyses simply as additional experimental households with different tax rates and disposable income. Control households are lumped into the sample to "increase estimation efficiency."[20] Investigators often perform some sort of pooling test to see if controls and experimentals can be lumped together. The question of whether controls are necessary in any fundamental sense except as low-cost observations has not been adequately addressed. One could also ask whether experimental observations are necessary and whether the existing experimental variations in prices and income are sufficient to estimate household response functions precisely.

A design procedure touted to be optimal must properly specify the objective function of the experiment as well as determine if the experiment is necessary. In the next section, we outline a decision-theoretic approach that provides a workable solution to this problem.

A Decision-Theoretic Approach

The key to developing a useful experimental design is a well-defined and meaningful objective function. Clearly defined objectives are critical in planning for a useful experiment. By forcing both the contracting agency and the investigator to specify clear and quantifiable objectives, it is possible to determine accurately whether an experiment is necessary and to produce a design that is able to discern treatment effects. Two main objectives were pursued in the negative income tax experiments: 1) computation of the net program cost of a national negative income tax program[21] and 2) estimation of the national labor supply response (work disincentive) to proposed negative income tax programs. The estimation of the labor supply response is a less ambitious goal than the costing of a national program and comes closest to the goal of most of the principal investigators.

It is also crucial that the results of a study be formulated in a way that can be effectively communicated to policymakers. We have found the research memoranda and papers on the negative income tax experiments to be very difficult to decipher even for readers with considerable econometric expertise. The results of statistical analyses are often reported without standard errors or evidence of diagnostic checking of any kind, without the number of observations in the estimation sample, and with poor labeling of tables and diagrams. More importantly, however, the extreme emphasis on point estimation and significance tests results often leads to misleading reports. For the policymaker interested in the costs of a national negative income tax, it is not sufficient to supply a point estimate of total costs. Some measure of the uncertainty in that point estimate due to estimation and possible specification error must also be supplied. It is extremely difficult to convey the uncertainty in point estimates just by supplying the estimate and the standard error of estimate. It is more useful to supply an interval and some probability statement about that interval. For a policymaker evaluating a negative income tax program, a statement such as "Given the information obtained in this experiment, we can say with a probability of .9 that the net program cost falls between A and B," is useful and understandable. Such a statement is one aspect of the predictive density of program costs, given the information available at the time of the report. The predictive probability density function expresses information about future costs on the basis of past sample and prior information with a due allowance for parameter uncertainty.

The policymaker can be supplied with various summary measures of the predictive density of costs, including the probabilities associated with various prediction intervals and measures of dispersion (variance, standard deviation and interquartile range) as well as plots of the density. In this way the policymaker can be assured that the accuracy of information available on key response parameters has been taken into account. Uncertainty with respect to the form of the response model can also be quantified, a problem treated at the end of this section.

The failure to report estimates of the precision of national cost estimates in the existing literature is all the more disturbing given that the national labor supply response may be very imprecisely estimated. To illustrate this point, let us examine the Keeley et al. (1978b) estimates of total labor supply response to various negative income tax programs. Keeley et al. estimate the hours response of heads of households and wives by applying a fitted labor response equation to a national probability sample of households derived from the Current Population Survey. By adding up the estimated hours supplied for each record in the file, Keeley et al. are able to produce national estimates of the work incentive/disincentive effect of various negative income tax programs. The labor supply response is figured in with other welfare and tax effects to impute the net cost of negative income tax programs. Their analysis depends critically on the quality of their estimated labor supply response functions.[22] Careful examination of their fitted response functions reveals that the fits are very poor and the coefficients measuring income and substitution effects are imprecisely estimated.[23] For example, in the equation for husbands, the standard error of regression is 720 hours per year. The mean number of hours worked per year before the negative income tax program is 1,999 for husbands. This suggests that the error variance in the labor supply response relationship is very large. (It also suggests that the model may be misspecified.) Even without taking account of estimation error, we would expect that predictions of supply responses to a national program would be very imprecise. In addition, the coefficients of the key variables are very imprecisely estimated. In the equation for husbands, not a single coefficient is estimated to within one significant digit of precision!

The results of microsimulation presented in table 7 of the Keeley et al. (1978b) paper do not include measures of precision. This gives the reader a false sense of the accuracy of these results. To illustrate the potentially enormous standard errors of prediction for these numbers, we will undertake some approximate standard error calculations. To obtain these crude figures, we must make many simplifying assumptions because we do not have access to the experimental and national data. However, the assumptions that are made bias downward our estimates of the standard error of prediction. Keeley et al. report an average reduction of 19 hours per year for husbands in a negative income tax program with support equal to 75 percent of the poverty level and a 50 percent tax rate. To put this figure in a form useful for computing program costs, we express all estimates as national totals. An average reduction of 19 hours implies a total reduction of 756 million hours per year for the total labor force of husbands. Keeley et al. have computed the total hours figure by summing up individual estimates as follows.

$$\hat{H}_{tot} = \sum_{i=1}^{N} \hat{H}_i \tag{1}$$

To derive the variance of the prediction error, $H_{tot} - \hat{H}_{tot}$, we must make some assumptions about the prediction errors for each individual's equation. We assume that the parameters of the response function are known and concentrate on the source of variability from the inherent randomness in the labor supply relationship, that is its error term. The calculation of the prediction error variance depends critically on the assumption used for the joint distribution of the labor supply of husbands. If it is assumed that the error terms of each husband are independent, the results are radically different from those obtained in the case in which even a small amount of dependence is allowed between units. To simplify the calculations, we assume that the $N \times N$ error covariance matrix has a simple patterned structure,

$$\Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \cdot & \cdot & \cdot & \cdot & \rho \\ \rho & 1 & \rho & \cdot & \cdot & \cdot & \rho \\ \cdot & \rho & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & & & \cdot & 1 & \rho \\ \rho & \rho & \cdot & \cdot & \cdot & \rho & 1 \end{bmatrix} = \sigma^2[(1-\rho)I_N + \rho \underset{\sim}{\iota}\underset{\sim}{\iota}'],$$

where $\underset{\sim}{\iota}$ is an $N \times 1$ vector with unit elements. Under these simplifying assumptions, the variance of the aggregate prediction error is given by,

$$\begin{aligned} \mathrm{Var}\,(H_{tot} - \hat{H}_{tot}) &= \sum_{i=1}^{N} \mathrm{Var}\,(H_i - \hat{H}_i) + \sum_{i \neq j} \mathrm{cov}\,(H_i - \hat{H}_i, H_j - \hat{H}_j) \\ &= N\sigma^2 + N(N-1)\,\rho\sigma^2 = N\sigma^2\,(1 - \rho + N\rho). \end{aligned}$$

The standard error, $\sigma = 720$ of the estimated tobit equation in table 3 of the Keeley et al. paper and an assumption about the value of ρ can be combined to produce a prediction error variance estimate. With $\rho = 0$, $N = 39.8$ million, and $\sigma = 720$, the prediction standard error is 4.54

million hours, which would produce a tight prediction interval around the point estimate of -756 million hours. If we allow for even a small amount of dependence by assuming $\rho = .01$, the prediction standard error increases to 2.86 billion hours. It should be emphasized that we are neglecting estimation error in these calculations by assuming that the conditional mean function is known to the analyst. The sensitivity of these calculations to what is assumed about the value of ρ is striking.

The fundamental goals of social experiments have been predictive. Either Bayesian or classical prediction techniques can be used to produce predictions and measures of precision. These same techniques can be used to determine the optimal design of experiments. Before the experiment is undertaken, some prior information is available on the key response parameters. In the case of labor supply, a number of studies with nonexperimental data[24] have produced both substitution and income elasticity estimates. These estimates can be combined to form a prior distribution for the response parameter vector, $p(\theta)$. The predictive density of hours can be computed using this prior[25] and compared to the predictive density which would be obtained after the experiment. The before and after predictive densities can be compared to determine if a given experiment has sufficient information value to warrant undertaking it. Peck and Richels (1986) use a similar decision-theoretic approach to indicate how to decide upon future research on the acid rain problem. Stafford (1985) also proposed a decision-theoretic approach for determining if negative income tax experiments are useful. Stafford proposed a social utility function and suggested that the information value of social experiments be measured via social utility. We avoid the problems associated with postulating a social welfare function and focus on the narrower goal of evaluating predictive accuracy.

Comparison of predictive densities can be accomplished by computing various scalar measures of differences in the distributions. Interval and probability estimates may be the most useful computations. For example, it may be that a 90 percent probability interval for the national costs of a negative income tax might include a wide range of values including cost greatly above and below current welfare costs. It may be the case that a given experimental design may sharpen up this interval to the point that policymakers may feel comfortable with the point cost estimate. One suspects that these types of calculations, when applied to the designs used in previous negative income tax studies, would suggest that the experiments had little informational value. Other summary measures that may be considered are variances and other moments and the divergence of two densities.

We have emphasized in earlier sections that uncertainty about model specification is one of the most serious problems confronting the design analyst. The frequent assumptions of log-normality and linear functional form of the labor supply function can easily be challenged. Problems with sample selection bias from the participation decision, attrition, and missing values also plague social experiments.[26] For example, Ferber and Hirsch (1979) point out that only 345 data points out of the more than 1300 enrolled households in the New Jersey experiment were actually used in estimating labor supply response. In the Seattle-Denver income maintenance experiments, approximately 1600 out of the total 2600 husband-wife households were used in fitting the labor supply equation. Useful optimal design procedures must consider the problem of optimal model selection and discrimination between alternative parametric models — see Box and Hill (1967), Covey - Crump and Silvey (1970), Guttman (1971) and O'Hagan (1978).

Given the considerable uncertainty regarding model specification, we are puzzled about the lack of discussion of predictive validation of models in the research reports of social experiments. The experimental data often contain numerous subsamples corresponding to different sites or different time periods or different treatment groups which could be used for validation purposes. For example, labor supply response functions fitted to Denver data could have been used to predict responses for the Seattle sample. If the labor supply function is well specified, the error terms should only contain random shifts due to tastes and other omitted characteristics of the households, and the prediction errors should follow the assumed error distribution in the model specification. If the response function cannot reliably extrapolate the results from one site to another, it is unreasonable to expect the same specification to be useful in predicting response to a national program.

A useful and easily generalized approach to model selection involves calculation of posterior probabilities of models in a Bayesian framework. Consider two different probability models for labor supply response, H, $p_1(H|\theta_1)$ and $p_2(H|\theta_2)$ where $\theta_1$and θ_2 are parameter vectors. For a given data set, we can compute the posterior probability of each model as follows:[27]

$$\Pr(\text{model i}|\text{data}) = \int l_i\,(\theta_i|\text{data})p_i(\theta_i)\Pr(\text{model i})d\theta_i/G \qquad (2)$$

where G = Pr(model 1|data) + Pr(model 2|data). The key ingredients in (2) are the likelihood function for each model, $\ell(.)$, the prior density for model parameters, $p_i(.)$, and the prior probability of the model. For comparison of models, we note that this approach does not require that the models be nested or that the models exhaust the set of plausible models. We do not adopt an accept/reject philosophy which eliminates models from future consideration even if the information in the data is insufficient to distinguish between the models. For many problems of practical interest, posterior model probabilities can be calculated without resort to asymptotic approximations.

One of the principal econometric problems encountered in the modeling of labor supply stems from the mass point at zero hours supplied in the empirical hours distribution. The analysts in the Seattle-Denver experiment used the truncated normal regression or tobit model to account for the massing of hours at zero:

$$\text{TOBIT model: } H_i^* = \underset{\sim}{X}_i' \underset{\sim}{\beta} + \epsilon_i \qquad \epsilon_i \sim \text{iid } N(0,\sigma^2)$$

$$H_i = \max\,(H_i^*, 0).$$

The tobit model can be considered as a special case of a two-equation model in which the first equation predicts labor force participation and the second equation gives labor supply response conditional on labor force participation.[28] A simple "two-part" model can be constructed by writing the density of H as consisting of a discrete and continuous part.[29]

$$p(H) = PR(H > 0)p(H|H > 0). \tag{3}$$

The participation equation can be a simple logit or probit binomial response model and the conditional distribution of hours for those in the labor force can be modeled with a simple regression function as follows:

$$\Pr(H_i > 0) = F(\underset{\sim}{x}_i'\underset{\sim}{\beta}) \tag{4}$$

$$\underset{\sim}{H} = \underset{\sim}{Z}\underset{\sim}{\theta} + \underset{\sim}{\epsilon} \quad \text{for } H_i > 0. \tag{5}$$

The distributional assumptions and parameter restrictions behind the tobit model are difficult to verify directly with data. One possible approach would be to fit the bivariate normal sample selection model of Heckman (1979) which nests the tobit specification. A significance test could then be performed on the parameter restrictions. However, the nested model hypothesis-testing approach often results in rejection of restricted models in favor of an unrestricted "super" model. Unfortunately, there is no clear connection between the tests of restrictions and the predictive performance of the models. It may well be that a tobit restrictions test may yield a rejection even though predictions from the tobit model may not differ much from a more complicated model. In fact, due to estimation error the simpler models may have a smaller out-of-sample mean squared error of prediction. Thus, predictive comparisons of alternative models should form the basis of model selection.

The normality assumption at the heart of the tobit and Heckman models is more difficult to check. The "two-part" model in equations (4) and (5) is more flexible and does not require strict normality assumptions. It would be of great interest to compare the tobit and two-part specifications. Equation (2) can be used to compute posterior model probabilities in much the same spirit as Rossi (1985). The integration over the model parameters necessary to compute the posterior probabilities can be performed with Monte Carlo numerical integration techniques as in Rossi (1985), or asymptotic normal expansions of the posterior distribution of the parameter vector can be used.[30] It should be noted that the posterior model probabilities can be used to average predictions from alternative models—see, for example, Geisel (1975, p. 229). We can hedge our bets on model specification by carrying along a small set of competing models until (if ever) one model specification dominates.

By stressing the predictive objectives of social experiments, it is possible to solve both optimal design and estimation problems in a manner consistent with a study's objectives. As a general rule, past analyses of social experiments have not stressed predictive validation or useful reporting of national cost and response estimates. We hope our suggestions will motivate a rethinking of the statistical methodology for use with experimental data.

Design and Analysis in a Dynamic Framework

Dynamic aspects of economic behavior have received increasing attention in both theoretical and empirical literature. Labor economists were among the first to stress the importance of dynamic economic models. Recent empirical labor economics often uses time series or panel data to explore dynamic econometric models. Much of this effort to understand economic dynamics was exerted during the time of the major social experiments. As Griliches[31] has pointed out, "theory . . . could be changing exogenously, thereby making the experiment less interesting than originally." Future experiments designed to gauge the response to social programs will have to be designed to illuminate some of the dynamic aspects of economic behavior.

In labor economics, dynamic models of labor supply have been developed to explain the life cycle pattern of labor/leisure allocations as well as the patterns in the spells of employment and unemployment. As many have pointed out, the labor supply schedule derived from a life cycle model of labor supply can differ substantially from the model based on a one-period demand-for-leisure analysis. Changes in income support programs can also affect the search for new jobs and the durations of periods of employment and unemployment. Much of the labor force is under implicit or explicit contracts which would be affected by changes in transfer programs. The stochastic process governing labor force participation is much more complicated than the simple Bernouli trial models behind common statistical specifications. The assumption that workers form rational expectations suggests that macro-level variables, which are useful in predicting the future course of output and wages, should be included in micro labor supply functions.

On a practical level, one of the most important dynamic aspects of social experiments is the problem of experiment duration. Most social experiments' test programs are intended to be implemented on a "permanent" basis or certainly for a much greater duration than the experiment. It seems obvious that households will react differently to permanent rather than transitory changes in wages and income. Moreover, experiments may be influenced by business cycle effects. The negative income tax experiments are of a short duration, typically lasting fewer than five years.[32] As an extreme case, the Housing Allowance Demand Experiment featured treatment cells in which households more often than not had to find new housing meeting rather arbitrary quality standards in order to qualify for small rent subsidies over only a two-year period. Other social experiments, notably some of the medical treatment experiments, have been conducted over much longer periods. In experimental situations in which treatment is limited to short durations, the challenge for the analyst is to calculate unobserved long-term effects.

As in our discussion of static models, we shall focus on the labor supply response relationship in the discussion of dynamic models. In the simple static model only current wages and income enter into the response function. Obviously, dynamic models enlarge this specification to include lagged response and input variables.[33] However, given the considerable debate over the appropriate dynamic theory of behavior, it may be wise to design in the context of an unrestricted transfer function model:

$$H_t = v_1(B)w_t + v_2(B)y_t + a_t$$

where w_t is the wage rate, y_t the income series, a_t follows a linear time series model, B is the backshift operator and $v_1(B)$ and $v_2(B)$ are rational functions of B. Of course, our ability to estimate lag structures can be severely limited by a short duration of an experiment. In such a situation, it is not clear that use of a static model will yield reliable results.

Time series analysts have developed a host of techniques for dealing with the adaptation of social and physical systems to environmental changes. In the intervention analysis pioneered by Box and Tiao (1975), time series models that allow for a wide variety of adjustment behavior are developed. In financial economics, critical events such as mergers or changes in government regulation are routinely studied with time series regressions and residual diagnostics in so-called "event" studies.

In order to design effective experiments for understanding complicated dynamic phenomena, we believe that a longitudinal design philosophy may be fruitful. In the labor supply problem, we observe very substantial individual variation coupled with complicated dynamic and program duration effects. The "one-shot" experiment may not be the most effective use of experimental resources. We propose a longitudinal design scheme in which an ongoing panel of households is used as the population for a large number of smaller-scale experiments. Many economists would agree that the lack of well-collected panel data on the household level is a critical problem for economic research. The few panel data studies available[34] have spurred tremendous interest and basic research in labor economics. It is our view that social investment in panel data collection on a permanent, ongoing basis has an extremely high social rate of return.

The existence of a reasonably large panel of households for which detailed information on most key aspects of household behavior is available radically reduces the start-up and overhead costs of experiments. It would not be necessary to perform huge screening operations to identify eligible experimental populations. A long time series of pre-experimental data would be available for each household so that investigators would not have to rely on recall. A national sample frame would be ensured at low cost. Numerous checks and mini-experiments could be performed to reduce measurement errors or, at least, to understand the properties of measurement error. It would not be necessary to train and organize a field interviewing staff from scratch as was done in the negative income tax experiments. The longitudinal and ongoing nature of such a project will also force those implementing the study to design carefully for the coordination between field operations and analytical database management, an area in which many unanticipated difficulties were experienced in the negative income tax experiments.

Many analysts have noted the tremendous individual variation in labor supply response. The poor fit of cross-sectional labor supply functions is usually attributed to large, unmeasurable individual taste effects. In a longitudinal design scheme, households could serve as controls for themselves by alternating experimental and control treatments. The diagram below indicates a possible design layout for longitudinal design.

	Period			
Household	1	2	3	4
1	O	O	O	O
2	X	O	X	X
3	O	X	X	O
4	X	X	X	X

X denotes treatment and O denotes no treatment. In this design, treatments are alternated for some households and remain fixed for other households. Thus, period 2 serves as a control period for household 2 in observing response in period 3, in addition to the usual "control" household number 1 which is never subjected to treatment. It is possible to subject different households to different durations of treatment to study the experimental duration effect. If these observations are spread over the business cycle, cyclical effects may be eliminated by averaging experimental response measurements over the business cycle.

As Sherwin Rosen has observed:

> We as a profession have engaged in excessive division of labor with regard to microdata collection. Thinking about survey instruments themselves and how they relate to economic phenomena and economic theories is probably an area where the social rate of return is fairly large.[35]

The sort of ongoing longitudinal data collection and experimental effort proposed here would encourage a wide range of research activities and give economists some private as well as social motivation for worrying about data collection and social experimentation.

Summary and Conclusions

We discussed some basic considerations involved in the evaluation of the methodology of social experiments. Many of the points raised seem obvious but, unfortunately, a number of them did not receive adequate attention in past social experiments in economics, for example in past negative income tax experiments. In our opinion, in most of these experiments, inadequate attention was given to formulating clear-cut attainable objectives. Feasibility studies and "test" or "pilot" experiments were nonexistent or not pursued vigorously enough. Serious measurement problems were encountered in these experiments and not dealt with adequately. Subject matter specialists, for example design statisticians, survey experts and outstanding subject matter theorists, were underrepresented or absent in the planning and execution of these experiments. Management and administration procedures were not completely satisfactory. The objectives of policymakers and of researchers usually were not clearly stated and in agreement. The experimental designs and the models on which they were based were inadequate in many cases. Last, the quality of reporting of results was generally far lower than could have been realized.

Some will say that the personal evaluations presented in the previous paragraph are "hypercritical" and that the negative income tax experiments constituted a valuable "learning experience." If so, this learning experience was very expensive and costly in terms of actual outlays and opportunity costs, including potential benefits associated with successful social experiments and other uses of scarce research resources. If learning was a main objective, then it is doubtful that the design actually used to achieve this objective was a very good one, as Rosen (1985, p. 137) has stressed.

In the previous sections, we have attempted to provide constructive suggestions for improving the methodology of social experiments. Among the points made, these seem particularly important:

1. It is critical to design experiments for successful prediction of observable outcomes that are central to the objectives of an experiment and to provide useful measures of predictive accuracy, preferably complete predictive distributions. Sample sizes should be large enough to yield needed precision in prediction and the range of the design space should be large enough to attain the objectives of an experiment.

2. When there is uncertainty regarding appropriate models for the observations, experimental designs that provide information for discriminating among candidate models should be employed. In this connection, it has been recognized that many existing designs are very sensitive to model misspecification, for example errors in choice of functional form, departures from independence, and use of univariate models when multivariate models are more appropriate.

3. A mixture of model-based and randomized designs seems most appropriate, with carefully designed randomization procedures employed to guard against certain types of possible model misspecification and prejudicial elements. ANOVA-based designs are not adequate because they are very sensitive to model misspecification, they involve the need for many experimental units when a large number of extraneous variables have to be controlled and, most importantly, they are incapable of generating the predictions required in many social experiments.

4. Predictive validation of models used in social experiments is essential. For example, the labor supply equations estimated in the Seattle experiments can be employed to predict labor supply using data from the Denver experiment and vice versa. Unsatisfactory predictive performance is usually an indication of model misspecification, differential selection and other types of bias, poor data, or other flaws. Further, vigorous diagnostic checking of models in other ways, for example residual analyses and outlier detection procedures, is also recommended. Use of inadequate models vitally affects the internal and external validity of experiments.

5. Use of point estimates alone to appraise costs of alternative negative income tax programs, in the very few cases in which cost estimates were derived, is inadequate. Measures of precision or predictive probability distributions should be provided and interpreted in easily understandable terms for the benefit of policymakers. For exam-

ple, the probability that the costs of a program lie between $20 billion and $30 billion, or the probability that the costs exceed $30 billion, can be calculated and reported. Similar remarks apply to predictions of changes in hours of work. In both of these instances, it is the case that departures from independence of outcomes for experimental units can have an extremely large impact on precision measures, for example standard errors of total estimated costs and changes in hours. There was little attention given to these points in past social experiments in spite of the fact that such dependencies are of great concern in survey work, econometric analyses of panel data and past work on experimental design.

6. Consideration of dynamic theoretical labor supply models leads to models for observations that are radically different from the generally static models employed in most past social experiments, and their use would lead to different designs for experiments and different models for analyzing experimental data. It is recognized that the forms of such dynamic models are often uncertain and thus the use of unrestricted transfer function models, univariate or multivariate, may be a good point of departure in design and analysis calculations. Also, as stated above, design for discriminating among models can be effective in dealing with model uncertainties in dynamic as well as static cases.

7. It is recommended that, when feasible, social experiments be linked to ongoing longitudinal data generating programs of well-established groups, a suggestion put forward years ago by Orcutt and Orcutt (1968). With such an arrangement, historical variables have been measured that are useful in before-and-after calculations, as is done in "event" or "intervention" analyses. A longitudinal design also permits individuals to be used as their own controls. This is a standard technique in experimental designs in biology and psychology.[36] Longitudinal designs can provide improved results and deserve much further study. In particular, their use permits exploration of dynamic models and possible successful extrapolation of experimental results to a national population, given that the longitudinal sample is a national one. Of course, administrative costs and other aspects of national, longitudinal experimentation require attention.

While we have pointed to many difficulties involved in past social experiments, it is our opinion that *properly conducted* social experiments can yield enormous social benefits. Perhaps the objectives of past experiments have been too broad and ambitious, a point also made by Griliches (1985). Limiting objectives of social experiments in economics may be essential for attaining success. Successful experience with experimentation in the areas of experimental economics, quality management, marketing, and agricultural economics tends to support this view. Also, "on-line" experimentation to appraise proposed changes in existing social programs probably will be fertile ground for social experimenters who draw on the growing quality-management literature on this topic.

Finally, we have noted that the negative income tax experiments were focused on variants of the negative income tax proposal put forward many years ago by Friedman (1962), Tobin et al. (1967) and others. Unfortunately, the information provided by these experiments was not generally considered in relation to possible fundamental modifications of the original proposals. Among possible modifications, one might be the use of time paths for tax payments different from those used in the negative income tax experiments. To subject a poor person who begins to work to a marginal tax rate of 50 to 70 percent *immediately* is an extreme "treatment." It seems feasible to formulate a more sensible temporal pattern of tax payments that would avoid these high, initial marginal rates, a topic for future research and, perhaps, additional social experimentation.

This research was financed in part by the National Science Foundation and the H.G.B. Alexander Endowment Fund, Graduate School of Business, University of Chicago.

[1]With respect to the New Jersey negative income tax experiment, Rossi and Lyall (1976) conclude, "When it came down to the congressional debate on FAP, it was evident that while the labor supply question interested some congressmen in a general way, concerns were addressed more to the total costs of a national program, an issue to which the experiment could not offer an answer even when complete. It is one of the apparent ironies of the experiment that while its motivation sprang from a strong concern with poverty and a desire on the part of both the experimenters and OEO to effect national welfare reform, its most substantial contributions may well be of a more scholarly sort in the area of experimental design and work behavioral response." (pp.176-77.)

[2]Rossi and Lyall (1976) remark with respect to the New Jersey experiment, "Economists played dominant roles in all phases of the experiment . . . Sociologists and social psychologists were to play minor roles in both the design and analysis. Not only are the strengths of economists reflected in the experiment, but also some of the mistakes and omissions of the experiment show the mark of the dominant economists." (pp. 10-11.)

[3]See Friedman (1962, p. 193) and Tobin, Pechman and Mieszkowski (1967) for calculations of the costs of existing welfare programs and of negative income tax programs.

[4]Spiegelman and Yaeger (1980, pp. 474-476) provide a useful discussion of reporting error in the Seattle-Denver income maintenance experiments. On the basis of a "large sample wage income study," they report that, "SIME/DIME participants reported between $100 and $300 less per year to the experiment than to the Internal Revenue Service. This amount is less than 5 percent of mean income. The variance in the amount underreported to SIME/DIME is on the order of $1,000, or about one-fifth of mean income. We observed that almost as many people overreported their incomes as underreported them." To understand these and other measurement problems, they state that "Further study of individual cases is necessary." These conclusions relate to wage income, which is probably easier to measure than non-wage income. Ferber and Hirsch (1982) present much useful material on measurement problems in the negative income tax experiments.

[5]It is surprising to us that M. Friedman (1986) and J. Tobin (1986), two leading experts on negative income tax proposals, did not play major roles in the experiments. Tobin (undated) did provide some comments on the design of the New Jersey experiment. He wrote, "I find an "anova" specification implausible for this problem. But I recognize that

there is a certain arbitrariness to any particular parametric specification.'' (p. 18)

[6]When single equation regression or response surface models were employed for design purposes, possible dependence of observations was rarely, if ever, considered.

[7]See McFadden (1985) and Ferber and Hirsch (1982) for valuable considerations of this range of issues.

[8]For example, observation of the population at various stages of the business cycle is relevant for negative income tax experiments. Seasonality is also relevant.

[9]See Hausman and Wise (1979, 1985, p. 208) and Robins and West (1986) for analyses of attrition bias and efforts to deal with it. Robins and West (1986) conclude on the basis of their analysis of Seattle-Denver data that ''Our results suggest that standard procedures of correcting for attrition bias do not always yield the proper results. The use of these procedures, however, depends to a large extent on the ability to model the attrition process and on the degree of attrition in the sample. In the SIME/DIME sample in which attrition was fairly modest . . . such techniques simply do not have the power to identify precisely the biases.'' (p. 337) In spite of these reservations, the authors conclude that ''attrition bias is not a serious enough problem in the SIME/DIME data to warrant extensive correction procedures.'' A similar conclusion was reached by Hausman and Wise (1979) in their analysis of the Gary income maintenance experiment. (p. 937). Ferber and Hirsch (1982, p. 75 and p. 95) have reservations about such conclusions, however.

[10]See the appendix to Hamilton et al. (1969) for a discussion of the importance of good management in large scale research projects.

[11]Rossi and Lyall (1976) give the following breakdown of total costs of the New Jersey negative income tax experiment.

A.	Administration and Research		
	Mathematica	$4,426,858	
	IRP-U. of Wisconsin	812,648	
	sub total		$5,239,506
B.	Transfer Payments		2,375,189
C.	Grand Total		$7,614,695

They state, ''The expenditures were a considerable overrun on the initial estimates of approximately $3 million. Most of the unanticipated expenses occurred on the research side. The handling of large and complicated data sets was simply much more costly than anyone anticipated.'' (p. 11) These comments underline the importance of good management techniques in the planning and execution of social experiments. For further discussion of these issues, see Ferber and Hirsch (1982).

[12]See footnote 1 for a possible illustration of this point in connection with the N. J. experiment.

[13]See Hausman and Wise (1985) for an exellent collection of articles on key aspects of analysis of economic experiments and Aigner and Morris (1979) for extensive discussion of designs of these experiments.

[14]See Watts and Bawden (1978) for discussion. Of course, some social experiments such as the Housing Supply Experiment are not feasible without a site approach. It would be impossible to discern a supply effect without involving a large percentage of households in particular housing markets.

[15]See Keeley et al. (1978a) for an example of this sort of calculation for a national negative income tax. Labor supply response functions were coupled with census household data in a ''micro-simulation'' of the national program.

[16]The experiences with changes in the New Jersey welfare laws during the course of the experiment highlight the importance of diversification. State-to-state differences in program implementation are to be expected in a national implementation of a negative income tax program. An experiment based on only one or two states cannot possibly take into account variation in local welfare programs.

[17]Conlisk and Watts (1979), p. 40.

[18]See Morris (1979) for a discussion of the finite selection model which essentially provides a randomization technique for providing more balanced experimental designs. The finite selection model utilizes the same sort of objective function and cost constraint as the Conlisk/Watts model.

[19]Hausman and Wise (1985) and Keeley et al. (1980) point out that current designs do not ensure that statistically significant treatment effects can be obtained.

[20]Keeley et al. (1978, p. 11).

[21]It was not until 1978 that Keeley et al. produced a thorough analysis of costs of a

national negative income tax based on labor supply response estimates from experimental data.

[22]See Keeley et al. (1978b, Table 3, p. 13) for estimates.

[23]The fitted model is a truncated normal regression or tobit model. The interpretation of σ as the standard deviation of prediction error conditional on knowing the model parameters is strictly not correct. σ should be interpreted as the standard deviation of prediction error for the latent variable. At levels of the independent variables for which little truncation occurs it is approximately correct to view σ as the root mean squared error of prediction.

[24]See Stafford (1985) for discussion of these studies and a table of elasticity estimates.

[25]Diffuse priors may be used in studies with little or unreliable nonexperimental data.

[26]See Hausman and Wise (1979a,b), Heckman (1978) and Hausman and Wise (1985) for discussion of modeling approaches to these problems.

[27]See Rossi (1985) for details of these calculations and an application to choice between alternative functional forms.

[28]Heckman (1976) makes this point.

[29]See Duan et al. (1984). This model has been employed earlier in econometrics by Orcutt, Goldberger and many others.

[30]See Zellner and Rossi (1984) for an example of this approach for binomial response models and Zellner (1971, 1984).

[31]Griliches (1985), p. 138.

[32]The Seattle-Denver income maintenance experiment contained treatments of three, five and 20 years. It is very difficult to find discussion of the results for 20-year treatments.

[33]Within the context of linear models, these dynamic specifications yield restricted transfer function models.

[34]The survey and income dynamics survey conducted at University of Michigan and the national longitudinal labor survey are examples.

[35]Rosen (1985, p. 137).

[36]Rossi and Lyall (1976), p. 42, fn. 24. See also Campbell and Stanley (1963), and Hall (1975).

References

Aigner, Dennis J. "A Brief Introduction to the Methodology of Optimal Experimental Design," *Journal of Econometrics*, 11 (1979a), pp. 7-26.

________. "Sample Design for Electricity Pricing Experiments: Anticipated Precision for a Time-of-Day Pricing Experiment," *Journal of Econometrics*, 11 (1979b), pp. 195-205

Aigner, Dennis J., and C.N. Morris, eds. *Experimental Design in Econometrics, Supplement to the Journal of Econometrics*, 11 (1979), pp. 1-205.

Box, G.E.P., and W.J. Hill. "Discrimination Among Mechanistic Models, *Technometrics*, 9 (1967), pp. 57-71.

Box, G.E.P., and G.C. Tiao. "Intervention Analysis with Applications to Economic and Environmental Applications," *Journal of the American Statistical Association*, 70 (1985), pp. 70-79.

Burtless, Gary, and Jerry A. Hausman. "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment," *Journal of Political Economy*, 86 (1978), pp. 1103-1130.

Burtless, Gary, and Larry L. Orr. "Are Classical Experiments Needed for Manpower Policy?" unpublished manuscript, 1986.

Cain, Glen G., and Harold W. Watts, eds. *Income Maintenance and Labor Supply*, Chicago: Markham, 1973.

Campbell, D.J., and J.C. Stanley. "Experimental and Quasi-Experimental Designs for Research on Teaching," in N.L. Gage, ed., *Handbook of Research on Teaching*, Chicago: Rand McNally & Co., 1963, reprinted in book form, *Experimental and Quasi-Experimental Designs for Research*, Chicago: Rand McNally & Co., 1973 (tenth printing).

Conlisk, John. "Choice of Response Functional Form in Designing Subsidy Experiments," *Econometrica*, 41 (1973), pp. 643-656.

________. "Design for Simultaneous Equations," *Journal of Econometrics*, 11 (1979), pp. 63-76.

________. "Comment", in Jerry A. Hausman and David A. Wise, eds., *Social Experimentation,* Chicago: U. of Chicago Press 1985, pp. 208-214.
Conlisk, John, and M. Kurz. "The Assignment Model of the Seattle and Denver Income Maintenance Experiments," Res. Mem. #15, Center for the Study of Welfare Policy, Stanford Research Institute, 1972.
Conlisk, John and Harold W. Watts. "A Model for Optimizing Experimental Designs for Estimating Response Surfaces," *Journal of Econometrics,* 11 (1979), pp. 27-42.
Covey-Crump, P.A.K., and S.D. Silvey. "Optimal Regression Designs With Previous Observations," *Biometrika,* 57 (1970), pp. 551-566.
Crutchfield, J.A., and Arnold Zellner. *Economic Aspects of the Pacific Halibut Fishery,* Washington, DC: U.S. Department of the Interior, Government Printing Office, 1963.
Duan, N., W.G. Manning, Jr., C.N. Morris, and J.P. Newhouse. "Choosing Between the Sample-Selection Model and the Multi-Part Model," *Journal of Business and Economic Statistics,* 3 (1984), pp. 283-289.
Ferber, Robert and Werner Z. Hirsch. "Social Experiments in Economics," *Journal of Econometrics,* 11 (1979), pp. 77-115.
________. *Social Experimentation and Economic Policy,* Cambridge: Cambridge U. Press, 1982.
Fienberg, S.E., B. Singer and J.M. Tanur. "Large-Scale Social Experimentation in the United States," in A.C. Aitkinson and S.E. Fienberg, eds., *A Celebration of Statistics: The ISI Centenary Volume,* New York: Springer-Verlag 1985, pp. 287-326.
Fisher, Ronald A. *Statistical Methods for Research Workers* (1st ed.), New York: Hafner Publishing Co., 1925.
Friedman, Milton. *Capitalism and Freedom,* Chicago: U. of Chicago Press, 1962.
________. personal communication, 1986.
Geisel, M.S. "Bayesian Comparisons of Simple Macroeconomic Models," in S.E. Fienberg and A. Zellner, eds., *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage,* Amsterdam: North-Holland Publishing Co., 1975, pp. 227-256.
Griliches, Zvi. "Comment," in Jerry A. Hausman and D.A. Wise, eds., *Social Experimentation,* Chicago: U. of Chicago Press, 1985, pp. 137-138.
Grossman, J.B. "Optimal Sample Designs With Preliminary Tests of Significance," *Journal of Business and Economic Statistics,* 4 (1986), pp. 171-176.
Guttman, I. "A Remark on the Optimal Regression Designs with Previous Observations of Covey-Crump and Silvey," *Biometrika,* 58 (1971), pp. 683-685.
Hall, Robert E. "Effects of the Experimental Negative Income Tax on Labor Supply," in Joseph A. Pechman and P. Michael Timpane, eds., *Work Incentives and Income Guarantees: The New Jersey Negative Income Tax Experiment,* Washington, DC: The Brookings Institution 1975, pp. 115-156.
Hamilton, H.R., S.E. Goldstone, J.W. Milliman, A.L. Pugh, E.R. Roberts, and A. Zellner. *Systems Simulation for Regional Analysis: An Application to River-Basin Planning,* Cambridge: MIT Press, 1969.
Hausman, Jerry A., and David A. Wise. "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," *Econometrica,* 47 (1979a), pp. 455-474.
________. "Social Experimentation, Truncated Distributions and Efficient Estimation," *Econometrica,* 45 (1979b), pp. 919-938.
________, eds. *Social Experimentation,* Chicago: U. of Chicago Press, 1985.
Heckman, James. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement,* 5 (1976), pp. 475-492.
________. "Sample Selection Bias as a Specification Error," *Econometrica,* 47 (1979), pp. 153-161.
Jeffreys, H. *Theory of Probability,* London: Oxford U. Press, 1967.
Keeley, Michael C. *Labor Supply and Public Policy: A Critical View,* New York: Academic Press, 1981.
Keeley, Michael C., Philip K. Robins, Robert G. Spiegelman, and Richard W. West. "The Estimation of Labor Supply Models Using Experimental Data," *American Economic Review,* 68 (1978a), pp. 873-887.
________. "The Labor Supply Effects and Costs of Alternative Negative Income Tax Programs," *Journal of Human Resources,* 13 (1978b), pp. 3-36.
Keeley, Michael C., and Philip K. Robins. "Experimental Design, The Conlisk-Watts Assignment Model, and the Proper Estimation of Behavioral Response," *Journal of Human Resources,* 15 (1980a), pp. 480-469.
________. "The Design of Social Experiments: A Critique of the Conlisk-Watts Model and Its Application to the Seattle-Denver Income Maintenance Experiments," in R.G.

Ehrenberg, ed., *Research in Labor Economics* (Vol. 3), Greenwich, CT: JAI Press, 1980b, pp. 293-333.
Kempthorne, O. *The Design and Analysis of Experiments,* New York: John Wiley & Sons, Inc., 1952.
MacCrae, E.C. "Optimal Experimental Design for Dynamic Economic Models," *Annals of Economics and Social Measurement,* 6 (1977), pp. 379-405.
McFadden, D.L. "Comment," in J.A. Hausman and D.A. Wise eds., *Social Experimentation,* Chicago: U. of Chicago Press, 1985, pp. 214-219.
Metcalf, Charles E. "Making Inferences from Controlled Income Maintenance Experiments," *American Economic Review,* 63 (June 1973), pp. 478-483.
________. "Predicting the Effects of Permanent Programs from a Limited Duration Experiment," *Journal of Human Resources,* 9 (1974), pp. 530-555.
Morris, C. "A Finite Selection Model for Experimental Design of the Health Insurance Study," *Journal of Econometrics,* 11 (1979), pp. 43-61.
O'Hagan, A. "Curve Fitting and Optimal Design for Prediction," *Journal of the Royal Statistical Society B,* 40 (1978), 1-42 (with discussion).
Orcutt, G.H., and A.G. Orcutt. "Incentive and Disincentive Experimentation for Income Maintenance Policy Purposes," *American Economic Review,* 58 (1968), pp. 754-772.
Palmer, John L. and Joseph A. Pechman, eds. *Welfare in Rural Areas: The North Carolina-Iowa Income Maintenance Experiment,* Washington, DC: The Brookings Institution, 1978.
Pechman, Joseph A., and P. Michael Timpane, eds. *Work Incentives and Income Guarantees: The New Jersey Negative Income Tax Experiment,* Washington, DC: The Brookings Institution, 1975.
Peck, S.C., and R.G. Richels. "On the Value of Information to the Acidic Deposition Debates," ms., Electric Power Research Institute, Palo Alto, CA, forthcoming in *Journal of Business and Economic Statistics.*
Rivlin, Alice. "Allocating Resources for Policy Research: How Can Experiments Be More Useful?" *American Economic Review Papers and Proceedings,* 64 (1974), pp. 346-354.
Robins, Philip K., and Richard W. West. "Labor Supply Response Over Time," *Journal of Human Resources,* 15 (1980), pp. 525-544.
________. "Sample Attrition and Labor Supply Response in Experimental Panel Data," *Journal of Business and Economic Statistics,* 4 (1986), pp. 329-338.
Rosen, S. "Comment," in J.A. Hausman and D.A. Wise, eds., *Social Experimentation,* Chicago: U. of Chicago Press, 1985, pp. 134-137.
Rossi, Peter E. "Comparison of Alternative Functional Forms in Production," *Journal of Econometrics,* 30 (1985), pp. 345-361.
Rossi, Peter H. "A Critical Review of the Analysis of Nonlabor Force Responses," in Joseph A. Pechman and P. Michael Timpane eds., *Work Incentives and Income Guarantees: The New Jersey Negative Income Tax Experiment,* Washington, DC: The Brookings Institution, 1975, pp. 157-182.
Rossi, Peter H. and K.C. Lyall. *Reforming Public Welfare: A Critique of the Negative Income Tax Experiment,* New York: Russell Sage Foundation, 1976.
Rubin, D.B. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology,* 66 (1974), pp. 688-701.
Spiegelman, Robert G., and K.E. Yaeger. "Overview," *Journal of Human Resources,* 15 (1980), pp. 463-479.
Stafford, F.P. "Income-Maintenance Policy and Work Effort: Learning from Experiments and Labor Market Studies," in Jerry A. Hausman and D.A. Wise, eds., *Social Experimentation,* Chicago: U. of Chicago Press, 1985, pp. 95-134.
Tobin, James. "Sample Design for NIT experiment," memo to Harold Watts and William Baumol, 19 pp, undated.
________. personal communication, 1986.
Tobin, James, Joseph A. Pechman, and Peter N. Miezkowski. "Is a Negative Income Tax Practical?" *Yale Law Journal,* 77 (1967), 1-27, reprinted in James Tobin, *Essays in Economics,* Cambridge, MA: MIT Press.
Watts, Harold W. and D.L. Bawden. "Issues and Lessons of Experimental Design," in John L. Palmer and Joseph A. Pechman, eds., *Welfare in Rural Areas: The North Carolina-Iowa Income Maintenance Experiment,* Washington, DC: The Brookings Institution, 1978.
Zellner, Arnold. *An Introduction to Bayesian Inference in Econometrics,* New York: John Wiley & Sons, Inc., 1971.
________. *Basic Issues in Econometrics,* Chicago: U. of Chicago Press, 1984.
Zellner, Arnold, and Peter E. Rossi. "Bayesian Analysis of Dichotomous Quantal Response Models," *Journal of Econometrics,* 25 (1984), pp. 365-393.

Discussion

*Jerry A. Hausman**

After a brief introduction Arnold Zellner and Peter E. Rossi turn to relevant considerations for evaluation of social experiment methodology. They discuss eight considerations which encompass design, management, and analysis of social experiments. In general their list provides a good common sense approach to the subject. I would like to stress their seventh point: that experiments should be designed to provide "accurate enough" predictions of various proposed policies along with measures of predictive precision. Because of the large amount of inherent variability in responses to tax and welfare policies, even within the assumption of a correctly specified model to evaluate the responses, two aspects of the Zellner-Rossi prescription should be emphasized. First, the range of policies that the experimental results will be used to evaluate must be specified with sufficient precision so that the experiment covers them. Otherwise, extrapolation outside the range of the experiment will be required, with undesirable consequences. This goal is often very difficult to achieve without increasing the costs greatly, and this aspect of design is especially dependent on the "specialists" Zellner and Rossi refer to. Second, the design and results must be able to supply results that are sufficiently precise to use. My greatest disappointment with the negative income tax experiments has been the low level of precision that arose from the results. Future social experiments should make sufficient precision in outcomes among their highest priorities in design.

Zellner and Rossi turn next to design considerations within a static framework. They criticize the Conlisk-Watts design for the negative income tax experiments for too restricted variation in experimental treat-

*Professor of Economics, Massachusetts Institute of Technology.

ment. Since I have discussed the Conlisk-Watts design elsewhere (Hausman 1982 and Hausman and Wise 1985), I will not return to previous ground. However, I would like to point out that the Zellner-Rossi criticism holds only within the context of a structural model of labor supply, for example the famous Elfving result for linear models, which places all the observations at the extreme points of the design space. In an ANOVA framework the response to each treatment point is estimated separately, so the Zellner-Rossi criticism does not apply. Even within the context of a structural model, I have considerable doubt whether I would want to use the responses. Our structural models are not usually sufficiently well specified that they can do a good job on extreme points in the sample space.

Next in their discussion of the Conlisk-Watts approach, Zellner and Rossi emphasize the specification of a demand equation for allocations of subjects across treatment groups. However, they fail to take into account the complexity of the actual demand equations that arise in response to government programs. For instance, the labor supply equation (leisure demand equation) will not be continuous even in the tax rates because of the nonconvexities in the budget sets. (See Burtless and Hausman 1978 and Hausman 1985.) The "housing gap" treatment in the Housing Allowance Demand Experiment has similar characteristics. This very complicated response surface is quite different from the response surfaces in many physical situations, where the responses are apt to be smooth. Zellner and Rossi should consider in more detail the complications for experimental design which these types of demand functions imply. These complications should induce a more favorable attitude to randomization procedures, which Zellner and Rossi discuss but do not strongly advocate.

Zellner and Rossi emphasize that a decision-theoretic approach would more likely lead to results that are usable by policymakers. While they stress a Bayesian approach to the problem of reporting results, I have found that an analogous "classical" approach, with point estimates and standard errors that account for parameter uncertainty, are straightforward to compute and seemingly well understood by public utility commission staffs who have evaluated results from experiments. I am in total agreement with Zellner and Rossi that the results of an experiment should be sufficiently precise to yield predictions with enough precision to give good guidance to policy. As Zellner and Rossi demonstrate in their analysis of the Keeley et al. results (1978) from the Seattle-Denver experiments, the negative income tax experimental designs do not lead to precise predictions about the labor supply response, which was certainly one of the major goals of the experiments. (Note that the Zellner-Rossi estimates of the prediction error of the Keeley results would be considerably larger if parameter uncertainty were accounted for, since this uncertainty is correlated across all observations in a microsimulation.)

Zellner and Rossi then turn to dynamic aspects of social experiments. They emphasize correctly that the experiments typically are of short duration, while the policies are permanent in nature and may therefore call forth a different response. However, I disagree with their suggestion that a Box-Jenkins times series approach would be a useful starting point for analysis. Lagged endogenous variables are quite difficult to treat in short panels because of initial condition problems; more importantly, the errors of measurement, which Zellner and Rossi emphasize earlier in their paper, have potentially devastating effects on times series type models of panel data. (See Griliches and Hausman 1986.) I do agree with their suggestions on the usefulness of panel data, which I discuss with respect to social experiments in Hausman (1982). However, it must be noted that panel data may raise the costs considerably for an experiment because of the necessity of keeping track of panel members. The cost trade-off between panel data and cross-section data would need to be considered, as Heckman has emphasized in recent research.

Zellner and Rossi conclude that the goal, design, execution, and analysis of the negative income tax experiments left much to be desired. I agree with these conclusions in large part. However, I believe their failings can be partly explained by the design and execution of the Gary and Seattle-Denver experiments before the lessons of the New Jersey experiment were learned. Presumably better experiments would be conducted now. My major point of disagreement lies in the analysis of the data: I believe that Zellner and Rossi have too much faith in structural models and that their time series approach to longitudinal data would not work well. But, we certainly agree that such experiments should be designed so as to be able to answer the important questions at issue in a precise enough manner to be useful for planning and policy purposes.

References

Burtless, Gary, and Jerry A. Hausman. "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment," *Journal of Political Economy*, 86 (1978), pp. 1103-1130.

Griliches, Zvi and Jerry A. Hausman. "Errors in Variables in Panel Data," *Journal of Econometrics*, 1986.

Hausman, Jerry A. "The Effect of Time on Economic Experiments," in W. Hildebrand, ed., *Advances in Econometrics*, Cambridge University Press, 1982.

________. "The Effect of Taxes on Labor Supply," in Alan J. Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*, 1986.

Hausman, Jerry A., and David A. Wise, eds., *Social Experimentation*, Chicago: University of Chicago Press, 1985.

Keeley, Michael C., Philip K. Robins, Robert G. Spiegelman, and Richard W. West. "The Estimation of Labor Supply Models Using Experimental Data," *American Economic Review*, 68 (1978a), pp. 873-887.

________. "The Labor Supply Effects and Costs of Alternative Negative Income Tax Programs," *Journal of Human Resources*, 13 (1978b), pp. 3-36.

Discussion

*Charles E. Metcalf**

Arnold Zellner and Peter E. Rossi review the conventional criticisms of the methodology of the early income maintenance experiments—which by now have had 19 years to accumulate—and provide their own suggestions for design of social experiments. Unfortunately, the authors' own recommendations fare poorly against the standards of criticism applied to previous work, and show insufficient evidence of the 19 years of experience that have accumulated since the design work for the first negative income tax experiment began. My comments follow the approximate sequence of the paper.

Considerations for Evaluating Methodology

In the first part of their paper, Zellner and Rossi offer eight considerations for evaluating the methodology of a social experiment. These conventional observations are largely correct but naively elaborated upon. For example, the call for "interaction between sponsors and bidders in the preparation of proposals" reflects a simplistic view of the competitive procurement process, but does touch on an important issue: the complex relations among contractor selection, project design, and project execution. Indeed, it is increasingly common for the design and execution phases of an experiment or evaluation to be the subjects of separate contract procurements.

Concerning the desirability of conducting a "pilot" trial of an experiment before proceeding with the "final" experiment, the distinc-

*President, Mathematica Policy Research, Inc. Views expressed are the sole responsibility of the author.

tion between these concepts is blurred in an evaluation environment cluttered with an extensive history of social experiments and demonstrations. We must also keep in mind that each "desirable" characteristic of an experiment has an opportunity cost, not the least of which is the passage of time. (Most suggestions for improving methodology tend to increase the duration of an experiment.) While many people—myself included—view the social experiments as having made important contributions to the policy process, provision of timely input with respect to originally specified experimental objectives is rarely one of them.

Failure to acknowledge opportunity costs also causes the authors to overstate another observation, which carries forward to their critique of the negative income tax experiments: "If the objectives . . . involve generalization . . . to an entire population, then the sample . . . *has to be* a sample from the relevant population." (Emphasis added.) It is equally true, however, that the program intervention being tested has to be the "relevant" intervention—in terms of features of program administration, duration, and so forth—and these two objectives are frequently in conflict. An experimental design stressing intervention with the right population is not clearly preferable to an experiment that restricts the population to improve the intervention.

Static Design Issues

Several static design issues raised by Zellner and Rossi are worthy of comment. First, the claim that the planners of the negative income tax experiments exaggerated the problems with administration and field operations of a national experiment is probably true for data collection, but *not*, in my judgment, for program administration. Recall that an effective implementation of a program intervention requires—aside from its placement in an effective evaluation structure—creation of a "relevant" program environment as viewed by the experimental subject; real and perceived independence of program administration from data collection; and at least some participation of welfare agencies in all jurisdictions covered by the experiment. These pressures all work to limit the number of jurisdictions covered by an experiment, and are further enhanced by the increasing prevalence of the view that "relevant" program interventions must be implemented by "real" program agencies rather than by experimenters, in order to be credible. This evolution is paralleled by a clear transition from experiments that test parameters to randomized demonstrations that test program interventions. There has been a recent trend toward the use of representative samples for demonstrations and/or experiments, but with cluster samples utilizing "real" program interventions in a relatively few sites.

Second, the authors criticize the negative income tax experiments for being too conservative in their choice of design parameters for the experiments. From a pure design perspective most experimenters would agree with the authors. But policymakers with whom the experimenters had to interact were reluctant to consider the concept of "extreme" experimental treatments outside the "policy-relevant range."

Third, the authors provide an extensive discussion contrasting the "response surface" and ANOVA approaches to design, and stressing importance of the analytic models that drive the experimental design. I agree with much of the authors' position here, and my disagreements with them are more often of form than of substance. Several points, however, are worth raising:

- The response surface approach is described as producing "nonrandomized" designs. This is true of the finite selection model extreme, but not of the Conlisk/Watts approach, which determines probabilities of selection for each element in the design space. So long as no probabilities of selection are permitted to go to zero, there exists an ANOVA equivalent for each response surface design.
- The potential damage caused by use of an inappropriate design model depends upon whether its use eliminates design points called for by the "correct" model, or whether it merely reduces estimation efficiency for the correct model. A linear or Cobb-Douglas model would spell disaster for the estimation of a translog function, but the converse is not so.
- In the (universal?) case where the correct model is not known with certainty, a risk-averse design strategy involves use of a model with more "dimensions" than specific models likely to be investigated, preferably with all probabilities of selection constrained to be positive. Inclusion of an ANOVA model as one of several weighted alternatives fulfills this objective. In such an environment it would not be surprising for the full design model never to be used for analysis.

Fourth, I do not regard the role of controls in social experiments as being "unusual" in their use of the status quo rather than the classical "no treatment" as the basis of comparison. The control group should reflect a relevant counterfactual, which may or may not meet the semantic definition of "no treatment." Consider also that removal from previously existing treatment is "no treatment" in only the most unrealistic of static worlds. As for whether controls are necessary except as cheap observations, this depends upon the experimental objective. For most policy purposes, as well as most reasonable predictive procedures, the relevant counterfactual is a critical component of evaluation. Indeed, I would regard the proper objective of the negative income tax experiments *not* to be estimation of the national cost of a negative income tax for comparison with external cost data for AFDC; rather, they should be providing internally valid *direct estimates* of the *differential cost*. I would argue this point on both policy and statistical grounds.

Fifth, I do not regard the discussion of cross-unit dependence as being particularly relevant from an empirical perspective, since rho = .01 is *massive* when applied for each unit to each of 40,000,000 other units, not "small" as alleged by the authors. If I were looking for a reason to disregard nominal standard errors obtained from the experiments, I would make a simple appeal to cluster sampling theory. For similar reasons I would not use labor supply functions fitted to Denver data to predict response for the Seattle sample, as suggested by the authors for validation, since the relevant sample size is *two* in too many dimensions. Rather, I would recommend a traditional split sample approach cutting across site boundaries to achieve that objective.

Dynamic Design Issues

The authors' discussion of dynamic design issues goes rather smoothly until they take seriously the notion of a longitudinal panel as the basis for drawing experimental samples, which takes the flawed concept of letting individuals be their own controls to an unfortunate extreme.

Concerning their general discussion, I would be careful to distinguish between two important but separate issues: the use of *limited-duration interventions* in place of relevant longer-term interventions (for example, the negative income tax experiments) and *limited-duration observation* of longer-term dynamic consequences. Time series models, for example, deal with the latter but not the former problem.

I have no quarrel with advocacy of better longitudinal data sources for continuing evaluation and research, often as an *alternative* to randomized experiments. The development of the SIPP panel appears to be especially promising. On the other hand, evidence is mounting that efforts to use longitudinal panels as comparison group alternatives to randomized control groups have been unsuccessful.

Similarly, the theoretical concept of an experimental panel has merit so long as it can provide an adequate sample, so long as the relevant program interventions can be applied to it, and so long as the sample points are disposable rather than reusable. Sample adequacy is a major problem, since many program interventions of policy interest are targeted to relatively small segments of the population. Earlier in my comments I questioned the ability to create the relevant program environment with a dispersed sample for most social programs of the sort earmarked for experiments or demonstrations.

Finally, the concept of reusing sample points in repeated experiments sounds fine when all interventions and impacts are static, but in a world of dynamic interventions and impacts the cross-experimental contamination effects would appear to destroy all credibility of the experimental results. Continuing panels for data collection are fine; for controlled interventions, extremely questionable.

Social Experimentation in the Face of Formidable Fables

*Dennis J. Coyle and Aaron Wildavsky**

The ways to prevent poverty are well known to us all. Allow every baby a chance to be born wanted, raised in good health, educated to full capacity, accepted upon individual merit, welcomed to a range of job choices according to capacity and interest, paid a good wage, insured at adequate levels against the economic hazards of the industrial economy, and assured a comfortable house in a supportive neighborhood, and opportunities for cultural enrichment, participation in the decisions affecting his own life, and survival into a respected and secure old age. In this age, these are not utopian goals. — Senator Fred Harris, 1970

Once upon a time there was a Little Red Hen who scratched about and uncovered some grains of wheat. She called her barnyard neighbors and said, "If we work together and plant this wheat, we will have some fine bread to eat. Who will help me plant the wheat?" "Not I," said the Cow. "Not I," said the Duck. "Guaranteed annual bread," said the Goose. "Then I will," said the Little Red Hen—and she did . . . She baked five loaves of fine bread and held them up for her neighbors to see. "I want some," said the Cow. "I want some," said the Duck. "I want some," said the Pig. "I demand my share," said the Goose. When the Farmer came to investigate the commotion he said, "You must not be greedy, Little Red Hen. Look at the oppressed Cow. Look at the underprivileged Pig. Look at the less fortunate Goose. You are guilty of making second class citizens of them . . . In other barnyards you would have to give all five loaves to the Farmer. Here you give four loaves to your suffering neighbors." And they lived happily ever after, including the Little Red Hen, who smiled and smiled and clucked "I am grateful, I am grateful." But her neighbors wondered why she never baked any more bread. — Merle Lofgren, 1970, *Corson County* (So. Dakota) *News*

*Doctoral candidate in political science and Research Assistant, Survey Research Center, and Professor of Political Science, respectively, University of California at Berkeley.

From the sober perspective of the 1980s, when grandiose spending schemes are few and budget deficits many, the pervasive confidence of the 1960s in the government's ability to cure social ills seems distant. But when the nation was enjoying an unprecedented level of affluence in the 1960s, no social goal was unattainable, at least in the minds of the social welfare specialists who made the pilgrimage to Washington during the Kennedy years. Professional expertise could be wedded to the bountiful American economy to erase the anomaly of the richest society in the world—the persistence of poverty. Nevertheless, the main manifestation of efforts to diminish poverty, the Family Assistance Plan, failed to win passage in Congress, and the New Jersey Graduated Work Incentive Experiment, designed to test the labor supply response, was part of that losing effort. The Family Assistance Plan was ultimately rejected, however, not because something was wrong with the research experiments that claimed to support it but because it failed to achieve the integration of political perspectives that would have made these claims acceptable.[1]

Setting the Stage: Policies in Search of Constituencies

"There was no 'demand' (in the interest group sense) for a war on poverty," writes Lawrence Friedman.[2] Rather, the "demand" came from the supply side, from the scholars who studied poverty and the bureaucrats who would conduct the battles. Self-interest was merged with social interest, in what Moynihan calls the "professionalization of reform."[3] The time appeared ripe for the rapid enactment of the Great Society poverty agenda. Even in a heady era of government expansion, however, the idea of a guaranteed income[4] was too extreme for a liberal President to embrace: "It will hurt its chances of ever being passed if it's connected with me," President Johnson warned.[5]

Proposals for a negative income tax appeared in the academic literature[6] and within the U. S. Department of Health, Education, and Welfare bureaucracy, gaining support from policy elites but generating little enthusiasm among politicians or the public. According to Robert Levine, the negative income tax "was not regarded as a serious proposal that could be enacted in less than a decade."[7] But interest in direct payments to the poor increased as disappointment with the training and service programs of the Great Society grew. For many liberals, the direct income transfer was just another strategy for bringing the fruits of affluence to the poor. Job training, housing, money—each program was a step toward liberating the individual from the shackles of impoverishment. "A psychology of scarcity produces anxiety, envy, egotism, (but) a psychology of abundance," which the guaranteed income would help achieve, according to Erich Fromm, "produces initiative, faith in life, solidarity."[8] For others, particularly economists, it was the market individualism of the income strategy that made it attractive.

The strongest and most consistent support for the negative income tax has come from economists. In May 1968, 1200 economists signed a petition in support of a negative income tax.[9] "It gives help in the form most useful to the individual, namely, cash," wrote Milton Friedman, and "It makes explicit the costs borne by society."[10] Liberal and conservative economists could agree that the negative income tax, which would provide universal, nationally uniform cash payments, would increase efficiency by reducing administrative costs and allowing individuals to pursue their own preferences. "The curse of the poor is literally their poverty. Give them more money," urged Paul Samuelson, "and not only they but their progeny can break through the vicious circle."[11] Guaranteed income was seen as a technically superior mechanism; a simple, efficient, visible transfer would replace a myriad of complex programs with their hidden costs. The devotion of economists to cost minimization through less bureaucratic procedures appears to have overridden the political implications of making it so easy to distribute welfare payments.

The negative income tax is an ideological hybrid, ill-fitted to common conceptions of liberalism or conservatism. Should conservatives oppose guaranteed income plans because they reward those who are irresponsible and indolent, or favor them because they allow poor individuals to pursue their own preferences free from government direction? Should liberals support income guarantees because they ease the misery of the unfortunate, or oppose them because they perpetuate inequality? The struggle over the guaranteed income can be described as a clash between three political cultures: hierarchy, libertarianism, and egalitarianism. These cultures are composed of people who share values justifying the social relations they deem desirable.

Hierarchs believe they have a social obligation to provide for the poor and that the poor cannot be trusted to look after their own interests. They will support paternalistic social policies, such as providing food, clothing and moral guidance to the needy. Egalitarians find paternalism offensive because it implies that some are wiser and should have more power than others. They will support poverty policies that seek to redistribute income and resources. Libertarian individualists believe it is the responsibility of each person to escape poverty and that the government should not tell anyone how to do it.

The Nixon Plan and the Welfare Establishment

Richard Nixon came to office seeking to show that a Republican could put together a domestic reform package that would meet the nation's social goals without overburdening the economy. He created an Urban Affairs Council, bringing in specialists in welfare policy who included several Democrats with experience in previous administrations, most prominently Daniel Patrick Moynihan. For policy proposals and data, the most available sources were bureaucrats within the Office of Economic Opportunity and the Department of Health, Education, and Welfare. The hibernation of the negative income tax plans during the Johnson years was coming to an end. "The welfare reform plan that had been brusquely dismissed by President Johnson was hauled out and dusted off," according to Martin Anderson. "The plan was rewritten, numbers were updated, and a few new ideas were added."[12]

Persuaded by Moynihan, Nixon reached into the welfare policy network and pulled out the Family Assistance Plan, which he presented to the public in a televised address on August 8, 1969. The Plan would have provided a minimum guaranteed annual income (subject to work requirements of debatable enforceability) of $1600 to every family in America; it also reduced benefits by 50 percent of earned income until the break-even point of $4000. Over the next 40 months, the Plan rose and died and rose again, as a series of bills won support in the House but wasted away in the Senate.

The Family Assistance Plan essentially was an attempt to provide enough enticement to each of the three political cultures to win their support, or at least to weaken their opposition. The Plan would extend benefits to millions of new recipients and establish a guaranteed income, which was supposed to please egalitarians. Limiting the plan to families would weaken egalitarian support but would be necessary to gain hierarchical backing by making it possible to argue that the Plan would strengthen a traditional institution—the family. Hierarchs would object to letting the poor control their own expenditures through direct payments, but this was necessary to entice individualists, who could support a program that would reduce government interference with the autonomy of poor people.

One might expect the Plan to have been supported by welfare workers and recipients. But this was not so. The criticisms from the welfare profession and welfare rights groups were that the guaranteed income was too low, was not universal, and was tied to work requirements. *Social Work,* the journal of the National Association of Social Workers, was filled with condemnations of welfare, and particularly the Family Assistance Plan, as a tool for the repression of poor people: "The welfare system . . . has . . . been used by society as an instrument to deny dignity to the soul. [Its features] contribute to totalitarian domination of the poor by those in power."[13]

Critics focused on the requirements that welfare recipents be available for suitable work or job training, or risk a reduction in their families' benefits, saying that the Plan would mean "compulsory work or starvation for mothers of school-age children."[14] Exceptions were to be made for mothers of young children, and child care and job training were to be provided. But the attempts to prepare recipients to be self-supporting were condemned as hoaxes: giving training to blacks on welfare would raise their aspirations, yet racial discrimination would prevent them from finding jobs. Ultimately, according to William Taylor, approval of Nixon's plan would be "social dynamite," causing a new wave of riots in the cities: the Plan "will . . . produce a new group of individuals who feel relatively deprived, frustrated, and angry enough to be susceptible to civil disorder and strife."[15]

Social welfare thinkers called for recognition of a national right to at least poverty-level subsistence for all individuals; they resented any suggestion that those who chose not to work were somehow inferior or should have their benefits reduced. "A person who does not work," complained Miriam Dinerman, "is virtually a nonperson."[16] "An individual should be able to claim maintenance as a legal right," George Hoshino argued, "unconditioned by the judgments of another person about his behavior."[17] The Nixon income plan, while aiding millions of new recipients, was seen as perpetuating too many of the evils of the welfare system to deserve support from the welfare establishment.

The years of the major congressional battles over guaranteed income plans, 1969-72, coincided with the zenith of the National Welfare Rights Organization (NWRO). The group was founded in 1966 under the direction of George Wiley, a former university chemist who resigned his faculty position to work directly for what he considered the rights of poor people. He was influenced by Richard Cloward and Frances Fox Piven, who then (they have since changed their views somewhat) conceived of welfare as a form of state oppression: the poor got a pittance in return for legitimizing the very institutional arrangements that kept them poor. Advocating a militant strategy of poor people demanding all the payments they were entitled to under aid to families with dependent children (AFDC), they argued this would overload the system and cause its collapse and replacement by a universal guaranteed income. A grassroots network was set up aimed at recruiting recipients of AFDC, the largest welfare program. In order to strengthen the organization, the National Welfare Rights Organization provided assistance not to the poor in general, but only to dues-paying members:

> Only members were provided with forms and assistance to obtain the special grants from the welfare department. As long as this information and assistance remained an exclusive payoff for members only—a "private good"—the NWRO membership and number of affiliate WRO groups multiplied. When these monetary incentives disappeared, or became widely available, the membership dropped almost as fast as it had risen.[18]

The special provisions embodied in welfare law were essential to the National Welfare Rights Organization because these made it worthwhile for clients to make special demands. Income maintenance would abolish or narrow special provisions.

The Family Assistance Plan would directly benefit the poor in states, mostly in the South, that provided payments below the minimum national level (originally $1600, later raised to $2400), but most National Welfare Rights Organization members lived in northern states that already had higher benefit levels. It would extend benefits to the working poor, who were also not represented in Organization membership. The Family Assistance Plan was in the best interests of the majority of the poor, according to Moynihan, but, "like the early trade unionists, the NWRO represented the aristocracy of welfare recipients."[19]

The Public View: A Different Definition of Poverty

In a Gallup poll conducted shortly after Nixon's August 1969 speech, 65 percent of respondents said they had a favorable opinion of the Family Assistance Plan, while only 20 percent said their opinion was unfavorable (table 1). The White House received over 2700 responses to the speech, characterizing 81 percent as favorable and only 9 percent as opposed. Media reaction was also positive; a Health, Education, and Welfare Department survey of newspaper editorials concluded 95 percent were favorable, and called newspapers in the 25 largest metropolitan areas "enthusiastic" about the plan.

A very different understanding of public attitudes, one that would predict opposition to any guaranteed income plan, emerges when respondents are asked to evaluate specific approaches to reducing poverty. Opposition to a guaranteed income, even for families, was strong and consistent (table 1). A Gallup poll in September 1965 found 67 percent opposed, 19 percent in favor. In May 1968, 58 percent were opposed, 36 percent in favor. The guaranteed income was opposed by every income group except those with incomes under $3000, who favored it by 3 percentage points, an insignificant margin. A December 1968 Gallup poll reported 32 percent favoring a guaranteed income and 62 percent opposed. Every income group was opposed.

Table 1
Public Opinion on Guaranteed Income Plans

In general, would you say you have a favorable or unfavorable opinion of Nixon's new welfare proposals? (Gallup: August 1969)

Favorable	65%
Unfavorable	20
No opinion	15

It has been proposed that instead of relief and welfare payments, the government should guarantee every family a minimum annual income. Do you favor or oppose this idea? (Gallup: September 1965)

Favor	19%
Oppose	67
No opinion	14

As you may know, there is talk about giving every family an income of at least $3200 a year, which would be the amount for a family of four. If the family earns less than this, the government would make up the difference. Would you favor or oppose such a plan? (Gallup: May 1968)

Total:		Income over $10,000:		Income under $3000:	
Favor	36%	Favor	26%	Favor	48%
Oppose	58	Oppose	68	Oppose	45
No opinion	6	No opinion	6	No opinion	7

Identical question (Gallup: December 1968)

Total:		Income over $10,000:		Income under $3000:	
Favor	32%	Favor	24%	Favor	43%
Oppose	62	Oppose	72	Oppose	44
No opinion	6	No opinion	4	No opinion	13

Opinions about welfare proposals shifted dramatically when the subject was guaranteeing or requiring jobs (which seem to have a similar meaning in the public's mind), not income (table 2). In 1964, 84 percent of the public agreed that men on relief who are physically able should be required to take any job offered that paid the going wage; support was identical among the lowest income group. According to a May 1968 Gallup poll, 78 percent of the public favored guaranteeing jobs to a wage earner in each family; support was very consistent, ranging from 75 percent among middle income recipients to 86 percent among blacks. A December 1968 poll yielded virtually identical results. In each case, a guaranteed income plan providing $3200 to a family of four was strongly opposed, yet a guaranteed jobs program providing the same income received wide support.

Table 2
Public Opinion on Work for the Poor

Here are some plans that have been suggested about the handling of relief. Will you tell me what you think about each one for this area? All men on relief who are physically able to work must take any job offered which pays the going wage. Would you favor or oppose this plan for this area? (Gallup: November 1964)

Total:		Lowest Income Group	
Favor	84%	Favor	85%
Oppose	11	Oppose	7
No opinion	5	No opinion	8

If men on relief, who are physically able to work, cannot find jobs, then they must work for the city on streets, parks, and the like. Would you favor or oppose this plan for this area? (Gallup: November 1964)

Total:		Lowest Income Group	
Favor	82%	Favor	79%
Oppose	12	Oppose	11
No opinion	6	No opinion	10

Another proposal (other than the guaranteed income) is to guarantee enough work so that each family that has an employable wage earner would be guaranteed enough work each week to give him a wage of about $60 a week or $3200 a year. Would you favor or oppose such a plan? (Gallup: May 1968)

Total:		Income over $10,000:		Income under $3000:	
Favor	78%	Favor	77%	Favor	83%
Oppose	18	Oppose	20	Oppose	16
No opinion	4	No opinion	3	No opinion	1

Identical question (Gallup: December 1968)

Total:		Income over $10,000:		Income under $3000:	
Favor	79%	Favor	76%	Favor	77%
Oppose	16	Oppose	22	Oppose	11
No opinion	5	No opinion	2	No opinion	12

The provision of goods and services in lieu of money was also supported by the public (table 3). In November 1964, 73 percent of respondents to a Gallup poll favored reducing the money given to persons on relief and giving them food and clothing instead; support was high even among the lowest income group. A 1969 poll found 68 percent of the public in favor of giving food stamps to families with earnings below $20 per week; agreement was strong within all income groups and regions.

Table 3
Public Opinion on Food and Clothing for the Poor

Some communities provide food and clothing to persons on relief, reducing the amount of money given to them. How do you feel about this? Do you think that it is a good idea or a poor idea? (Gallup: November 1964)

Total:		Lowest Income Group:	
Good idea	73%	Good idea	65%
Poor idea	19	Poor idea	25
No opinion	8	No opinion	10

A United States senator has proposed that the Government give free food stamps to all families whose earnings are under twenty dollars a week. Do you favor or oppose such a proposal? (Gallup: April 1969)

Total:		Income over $10,000:		Income under $3000:	
Favor	68%	Favor	67%	Favor	76%
Oppose	25	Oppose	27	Oppose	17
No opinion	7	No opinion	6	No opinion	7

Would you favor or oppose giving food stamps at a greatly reduced rate to those whose earnings are twenty to sixty dollars a week? (Gallup: April 1969)

Favor	60%
Oppose	31
No opinion	9

The public shared with policy elites a concern for the poor and a belief that the government should do something to alleviate poverty. Important aspects of the negative income tax received public support: that a minimum for poor families should be guaranteed by government, and that the working poor should receive benefits. But the mass public opposed a guaranteed income, preferring instead to guarantee, and require, work. "Not many Americans outside the antipoverty community," writes Hugh Heclo, "seemed to accept the concept of a right to income as such but only to the necessities income might buy."[20] Representative Landrum voiced this sentiment when he complained, "The priorities of this bill [the Family Assistance Plan] are wrong. They are: cash, first; food, second; and work, third. I believe there should be a reversal in priorities: work, first; food, second; and cash, last."[21] This partly explains why there was to be a food stamp program but not an income maintenance program.

Members of the public prefer a different solution—jobs, not money—because they have a different definition of the problem. If poverty is the lack of money, then the provision of money should end poverty. But if poverty is the lack of a job, and the discipline and self-respect that go with it, then transferring money may only gloss over the poverty problem. Martin Anderson puts it bluntly when he writes:

> The provision of an adequate income may eliminate poverty in the official sense, but it does not guarantee that those who receive welfare will spend that income in a manner that also eliminates the characteristics that many people associate with poverty . . . If they personally value nice cars, good liquor, and gambling, they may not have much money left for housing, clothing and food.[22]

Poverty, for the public, is not only a lack of resources but also a problem of behavior.

The public's attitude toward poverty is that giving money to those who cannot handle it is futile. Better to follow a paternalistic policy of giving the poor what is good for them—such as food and clothing—and requiring and guaranteeing work, which will give them the moral character to be self-reliant. Then (and only then) should they receive the reward—the freedom to spend their earnings as they please. Rewards should flow from taking advantage of opportunities, not from getting rewards in order to seek opportunities.

The public, then, had an ambivalent attitude toward the reforms embodied in the Family Assistance Plan. They supported the President's determination to "clean up the welfare mess," but distrusted transfer payment programs for the poor. The Family Assistance Plan that failed was no more antithetical to the general public than was tax reform at a time when majorities thought dealing with the deficit was more important. In each case it was political elites, not the public, who took the lead; it is to their activities that we now turn.

The Battle in Congress: Clash of the Fables

Guaranteed income plans, such as the Family Assistance Plan, were caught between two opposing cultural myths: One, that the ways to end poverty, in the words of Senator Harris, "are well known to us all," and thus any plan that does not provide for the immediate lifting of every individual out of poverty is inadequate. In this egalitarian view, the costs of an aggressive program to the social values or economic resources of the society are inconsequential, or cannot ethically be considered. The opposing myth, as captured in the tale of the Little Red Hen, is that any assistance to the poor that is not strongly tied to individual work effort will destroy the moral fabric and bankrupt the society.

When Cavala and Wildavsky asked members of Congress about the guaranteed income prior to Nixon's proposals, they found widespread, automatic support among "safe-seat" liberals. "They knew only that guaranteed income was a liberal issue and that they were liberals; ergo, their support was automatic . . . There was little concern with the moral or even the knotty technical issues involved in guaranteed income."[23] Liberal support was sufficiently strong in the House to aid passage of Nixon's guaranteed income plan, but in the Senate the coalition came unraveled.

The guaranteed income plans died in part because liberals, encouraged by the Administration's portrayal, began believing their exaggerated rhetoric about the inadequacy of the Nixon proposal. In each house, liberals proposed more egalitarian alternatives to the Family Assistance Plan that would broaden coverage to include all individuals and greatly raise the minimum floor. "Minimal financial security should be a right of citizenship," asserted Senator George McGovern, who promoted a Human Security Plan that would guarantee jobs and income to all. Representative William Ryan introduced the Income Maintenance Act, a more generous plan based on an Office of Economic Opportunity draft, and also cosponsored the National Living Income Program, an outline of which had been drafted by the economist James Tobin. But the Nixon strategy of downplaying the egalitarian nature of the welfare reform proposals carried over into some of the congressional alternatives. The sponsors of the National Living Income Program discovered that "the President's plan appears to be limited to families with children."[24] There was "no justification for such discrimination," according to Representative Goodell.[25] Their own plan, which emphasized that payments would go only to families (a buzzword pleasing to hierarchists), decreed in the fine print that a family shall consist of "at least one claimant." If only aid to families was politically feasible, then family would have to be redefined so that none would be excluded.

Family stood for something more than a legal definition of people living together or related to one another. Family symbolized social order. Its children stood for the deserving poor, the dependent people who could not be expected to work until society helped them to help themselves. The sacrificial ethic of the hierarchical collective, in which the better off help the worst off, just as officers go first in battle, made welfare into a matter of mutual obligation, the one to give in good grace, the other to receive in gratitude, saying metaphorically that it is a good system that takes care of its own. Eliminate family and you wipe out Moynihan's carefully crafted effort to blunt the usual attacks on welfare (the good us versus the feckless them) through an integrative solution. For if Americans were all part of the same family, they were just helping themselves.

Statements by Moynihan, Patricelli and others that the welfare reform plans were major steps toward a universal guaranteed incom may have failed to impress egalitarians, but they did not go unnotice by individualists. The Family Assistance Plan was condemned b Representative John Rarick, who quoted a column in the *Economi Council Letter* calling the guaranteed income a "scheme for legal plunde on a scale without precedent in all history."[26] For those who claimed th guaranteed income would extend the legacy of the New Deal, *The Wal Street Journal* quoted the wisdom of Franklin Roosevelt, who proclaime in his 1935 State of the Union address that "To dole out relief is t administer a narcotic, a subtle destroyer of the human spirit."[27]

The two camps also made different empirical assertions. Neithe side seemed terribly concerned about marshaling evidence to support it arguments. Liberals took it as a matter of faith that the guaranteed in come would not do significant harm to the economy; to question tha would be to criticize the character of the poor. Conservatives believe that only work, not money, could end poverty. " 'If you cut your ow wood,' the philosophy of the conservative legislators went, 'it warm you twice.' "[28]

The integrative solution embodied in the Family Assistance Plan—family support for conservatives,[29] extension of benefits for liber tarians, and reduced bureaucracy and greater autonomy fo liberals—failed because adherents of these ideologies were not persuad ed there was enough in the Plan for them.

The Income Maintenance Experiments: Policy Analysis as Political Ammunition

Social scientists often cringe when they see how their research is distorted, if it is noticed at all, in the political arena. Once empirical studies are disseminated, political expediency may overwhelm the search for truth, or so the common wisdom goes. "Policymakers, while not totally subjective and nonrational, will use whatever data are at hand to support their case," writes Ernest Stromsdorfer. "Canons of evidence are not ignored but are selectively applied."[30] In assessing the use of analyses of Great Society programs, Henry Aaron concludes that "Evaluation was a political instrument to be trotted out when it supported one's objectives or undercut one's opponents, and to be suppressed, if possible, when it opposed one's objectives or strengthened one's opponents."[31]

What started as an experiment intended by the Office of Economic Opportunity to strengthen the case for a guaranteed income several years in the future soon became of immediate political relevance when

Nixon proposed welfare reform along the lines of the New Jersey experiment. Positive findings would have to appear early if they were to aid passage of the bill. "Well before [the experiment] was completed," recalls Moynihan, "a President had embraced its principles and hoped-for conclusions . . . Inevitably, there arose a conflict between the methodological demands of social science and the political needs of Congress and the Administration, and perhaps just as inevitably, the latter won out."[32] Officials "broke into" the data and produced their "preliminary" findings in February 1970.

"There is no evidence that work effort declined among those receiving income support payments," the Office of Economic Opportunity report concluded. "On the contrary, there is an indication that the work effort of participants receiving payments increased relative to the work effort of those not receiving payments."[33] These findings ran counter to the predictions of economic theory that income supplements would encourage people to work less, but were welcome news to those supporting the bill.

Officials later backpedaled a bit, but the initial findings were cited repeatedly by politicians and economists who supported the negative income tax. A 1971 report on the New Jersey experiment again concluded that work effort is "undiminished by negative tax transfers."[34] Andrew Brimmer, an economist and member of the Board of Governors of the Federal Reserve System, told an audience in June 1971 that "There is well-founded evidence [e.g., the results of the New Jersey Graduated Work Incentive Experiment] showing that poor people prefer to work—even when they receive an income supplement."[35] Moderates and liberals in Congress used the findings to support guaranteed income plans. A universal income floor would yield "great economic benefits [and] create incentives [to work. The New Jersey experiment] shows this very clearly," Senator Harris claimed.[36] When asked on the House floor whether the Family Assistance Plan would create incentives to work, Ways and Means Committee Chairman Wilbur Mills replied that, as regards New Jersey, "their final report will indicate the success of that experiment."[37] References to the experimental results in the congressional hearings and floor debates are few; by and large, the negative income tax experiment was ignored, but, when it was cited, the misleading preliminary findings received the most attention.

After the Government Accounting Office criticized its preliminary findings, the Office of Economic Opportunity backed away from the shaky claim that the income guarantee actually increased work effort.[38] Now officials asserted that the most reasonable conclusion from New Jersey was that work effort did decrease but not by much. "We have not picked up any precipitous decline in work effort. That is the major crux of our findings," John Wilson, assistant director of the Office of

Economic Opportunity, reported to the Senate Finance Committee. He emphasized the positive by claiming that "low-income people are strongly work motivated," basing that assertion not on behavioral evidence, but on an opinion survey of recipients.[39] The New Jersey results, William Morrill of Health, Education, and Welfare concluded:

> . . . clearly indicate that a negative tax type plan . . . will not trigger *large* scale reductions in work effort . . . only *minor* cost implications should be expected . . . Offsetting these would be the potential for substantially reducing income poverty, increasing the command of the poor over material goods and services, and enhancing their freedom to choose among economic options (emphasis added).[40]

The New Jersey team apparently followed the advice that if you cannot win, declare victory. They had found evidence that income guarantees could decrease, not increase, work effort; hence they concluded that "the burden of proof would now appear to be on those who assert that income maintenance programs for intact families will have very large effects on labor supply."[41] According to those sharing this view, the battle for the negative income tax was nearly won: "Public opposition to coverage of all intact families by an income-related cash-transfer program . . . should decrease," claimed Michael Barth, Larry Orr and John Palmer. "The case for a work test . . . is weakened."[42]

Ultimately, the New Jersey experiment had little to do with the political fortunes of the Family Assistance Plan. But the Office of Economic Opportunity cannot be faulted for lack of trying. In its hands, through creative interpretation of results, the negative income tax experiment was partially molded into arguments for the proponents. But it was still a policy experiment, with ambiguous and undramatic results, hardly equal to the task of overcoming fundamental cultural disputes, the gulf between the egalitarian nirvana of Senator Harris and the libertarian lesson of the Red Hen fable.

The Modest Role of Experimentation

Policy research has been criticized as being an impediment to reform. Henry Aaron says that analysis is "profoundly conservative," strengthening opposition to change by pointing out the imperfections in any reform proposal. David Greenberg and Philip Robins claim that "The probability of enactment [of proposals such as the guaranteed income] was reduced as a direct consequence of experimentally testing them."[43] This conclusion is easy to reach if one believes the policies advocated are innately good and would be supported by the public if they properly understood them. When enlightened policies backed by the President and key members of Congress fail, a sinister force must be at work, and the misuse of policy analysis by the opponents of change becomes a convenient scapegoat.

If the only choices are to maintain the status quo or completely transform the society, then perhaps experimentation would have a conservative role. But this is hardly the usual picture; more often policy research has a moderating influence, showing us that a change in policy would neither be as beneficial as some might hope nor as harmful as others might fear. This, indeed, is the legacy of the income maintenance experiments. They showed that there would be work reductions, but of modest proportions; they also demonstrated that it was practical to administer the program. The consequences of the experiments were conservative only in light of inflated promises: that the income maintenance plans were not guaranteed incomes, would increase work effort, and would reduce the burden on taxpayers. No amount of research could credibly support these claims.

The guaranteed income proposals failed not because policy research had a conservative effect, but because they were orphaned by the welfare establishment and its egalitarian supporters. Caught between conflicting demands, the Family Assistance Plan was seen as not giving enough to each to secure passage. This insufficiency was a consequence of the legislation itself and the expectations of the times; failure cannot be blamed on the experiments. "Research, no matter how relevant and competent," Michael Barth and his coauthors remind us, "cannot tell us what national policy ought to be."[44]

Far from preceding policy, data are inextricably intertwined with the theories on which public policies are based. In formulating policy, therefore, there is no unalterable need to get the numbers straight before doing anything else. On the contrary, it is the policy one has in mind that determines which data, accurate to what degree, are relevant.[45]

Scientists acknowledge the "objectivity" of results by certifying among themselves the integrity of the process, not by direct apprehension of the facts. Appeal to the facts to resolve disputes is possible only when there is prior consensus, both as to the implicit conceptual framework (the language of discourse) and the rules of resolution. And this consensus was lacking at the time of the New Jersey negative income tax experiment.

Why did this integrative approach, blessed with an integrative name—family assistance—and an integrative argument—use market methods to secure equity and social order—fail in garnering sufficient support? The Family Assistance Plan died not because the demise of welfare reform was inevitable but because in those days the elites who spoke for egalitarianism would not go along. In their eyes, reform was certainly too little, perhaps too late. Everyone had to be made better off.

No means test was permissible. Nothing could be left for tomorrow. The existing system was so rotten that only the most radical change was tolerable. They demanded far greater expenditure so that all welfare recipients would immediately receive substantially more, while denigrating the considerable change that could be accomplished, thus casting a pall over income maintenance before it was defeated.

Yet facets of the negative income tax have been incorporated into poverty policy. Although the proposal of a guaranteed income for families died, a guaranteed income for the needy aged, blind and disabled was enacted (Supplemental Security Income). A universal income was rejected, yet a far-reaching food stamp program was adopted. These alternatives succeeded because they were perceived as more integrative solutions: SSI provided aid to a group that even libertarians might concede merited special assistance; food stamps pleased hierarchists by supporting an important institution—the farm—while providing for a basic need—food for the unfortunate.

What could have happened, we ask, in the spirit of counterfactual history, if the Family Assistance Plan, like the 1986 tax reform legislation, had been accepted by egalitarians as a basis for negotiation rather than rejected as inherently flawed? The morale of the sponsors would have received a tremendous boost. The prospects of gaining credit for an historical change, rather than taking the blame for the failures of welfare, might have engendered a broader appeal. As long as the presumed beneficiaries thought it good for them, the prospects for social peace, dear to the hearts of hierarchs, and for self-reliance, as individualists desired, might have appeared brighter.

Had the Family Assistance Plan passed, the New Jersey negative income tax experiment might have been hailed as a visionary social experiment, policy research at its best, providing it was practical to implement a novel social solution. But its success would have been due far more to facilitating political circumstances that allowed for ideological compromise than to any consequences of experimentation.

The role of the income maintenance experiments in the political battles over the Family Assistance Plan paled in comparison with the vigorous ideological clashes. "Rarely, if ever," Burke and Burke tell us in their authoritative book, "has a proposal met with such misinformed but energetic attack."[46] Income maintenance challenged fundamental beliefs about the good life—how to live it, who is obligated to whom for what—and it was on this ground that the battle was fought. Experimentation may point the way toward specific policy solutions once there is sufficient consensus to make broad support possible, but research cannot replace the dialogue among supporters of different ways of life. Were research a substitute for mutual persuasion, there would be no democracy, no pluralism, and, in the end, no decent research.

[1]See also Dennis J. Coyle and Aaron Wildavsky, "Requisites of Radical Reform: Income Maintenance Versus Tax Preferences," *Journal of Policy Analysis and Management*, May 1987.

[2]Lawrence M. Friedman, "The Social and Political Context of the War on Poverty: An Overview," in Robert H. Haveman, ed., *A Decade of Federal Antipoverty Programs*, New York: Academic Press, 1977, p. 35.

[3]Daniel Patrick Moynihan, "The Professionalization of Reform," *The Public Interest*, Summer 1965, p. 6.

[4]We will use the terms guaranteed income and negative income tax interchangeably as forms of income maintenance, to describe programs that would provide a minimum income for eligible recipients.

[5]Robert Harris, Review of Henry J. Aaron, *Politics and the Professors: The Great Society in Perspective* in *Knowledge*, 2 (1981), p. 455.

[6]Milton Friedman, *Capitalism and Freedom*, Univ. of Chicago Press, 1962; Robert Theobald, ed., *The Guaranteed Income: Next Step in Economic Evolution?*, Garden City, N.J.: Doubleday, 1965; James Tobin, "Improving the Economic Status of the Negro," *Daedalus*, Fall 1965, p. 878; James Tobin, "The Case for an Income Guarantee," *The Public Interest*, Summer 1966, p. 37; C. Green and R. J. Lampman, "Schemes for Transferring Income to the Poor," *Industrial Relations*, 6 (February 1967), p. 121; and Joseph Pechman, James Tobin, and Peter Mieszkowski, "Is a Negative Income Tax Practical?" *Yale Law Journal*, 77 (1967), p. 1.

[7]Robert A. Levine, "How and Why the Experiment Came About," in Joseph A. Pechman and P. Michael Timpane, eds., *Work Incentives and Income Guarantees: The New Jersey Negative Income Tax Experiment*, Washington, D.C.: The Brookings Institution, 1975.

[8]Erich Fromm, "The Psychological Aspects of the Guaranteed Income," in Theobald, *The Guaranteed Income*, p. 176.

[9]*Congressional Record*, August 4, 1969, p. 22187.

[10]Friedman, *Capitalism and Freedom*, p. 192.

[11]*Congressional Record*, August 4, 1969, p. 22188. (This in the days before the spectacle of suddenly oil-rich countries raised doubt whether the difference between rich and poor was mostly money.)

[12]Martin Anderson, *Welfare: The Political Economy of Welfare Reform in the United States*, Stanford, Calif.: Hoover Institution Press, 1978, p. 8.

[13]Betty Mandell, "Welfare and Totalitarianism," *Social Welfare*, 16 (January 1971), p. 17.

[14]Alfred J. Kahn, "Guaranteed Protestant Ethic," *Social Work*, 15 (January 1970), p. 3.

[15]William H. Taylor, "Unintended Consequences of the Nixon Welfare Plan," *Social Work*, 15 (October 1970), p. 15.

[16]Miriam Dinerman, "Mr. Harris and the Elephant Game," *Social Work*, 15 (January 1970), p. 114.

[17]George Hoshino, "A Conceptual Analysis of the Nixon Welfare Proposal," *Social Casework*, 51 (March 1970) p. 157.

[18]Guida West, *The National Welfare Rights Movement: The Social Protest of Poor Women*, New York: Praeger, 1981, p. 312.

[19]Daniel P. Moynihan, *The Politics of a Guaranteed Income: The Nixon Administration and the Family Assistance Plan*, New York: Random House, 1973, pp. 334.

[20]Hugh Heclo, "The Political Foundations of Antipoverty Policy," in Sheldon Danziger and Daniel Weinberg, eds., *Fighting Poverty: What Works?* Cambridge, Mass.: Harvard Univ. Press, 1986, p. 326.

[21]*Congressional Record*, March 24, 1970, p. 9107.

[22]Anderson, *Welfare*, p. 34.

[23]Bill Cavala and Aaron Wildavsky, "The Political Feasibility of Income by Right," *Public Policy*, 18 (Spring 1970), p. 321.

[24]*Congressional Record*, September 12, 1969, p. 23396.

[25]*Congressional Record*, March 31, 1970, p. 9841.

[26]*Congressional Record*, April 27, 1970, p. 13328.

[27]*Congressional Record*, January 30, 1969, p. 2288.

[28]Cavala and Wildavsky, "Political Feasibility of Income by Right."

[29]This argument would be weakened several years later by evidence from income maintenance experiments in Seattle and Denver that transfer payments increased family breakup, but by then the Family Assistance Plan was long gone.

[30]Ernest W. Stromsdorfer, "Social Science Analysis and the Formulation of Public Policy," in Jerry A. Hausman and David A. Wise, eds., *Social Experimentation,* Chicago: University of Chicago Press, 1985, p. 258.

[31]Henry J. Aaron, *Politics and the Professors: Studies in Social Economics,* Washington, D.C.: The Brookings Institution, 1978, p. 32.

[32]Moynihan, *Politics of a Guaranteed Income,* p. 191.

[33]*Congressional Record,* December 31, 1970, p. 44505.

[34]David Elesh, Jack Ladinsky, Myron J. Lefcovitz and Seymour Spilerman, "The New Jersey-Pennsylvania Experiment: A Field Study in Negative Taxation," in Larry L. Orr, Robinson G. Hollister and Myron J. Lefcovitz, eds., *Income Maintenance: Interdisciplinary Approaches to Research,* Chicago: Markham, 1971, p. 29.

[35]Moynihan, *Politics of a Guaranteed Income,* p. 344.

[36]*Congressional Record,* February 10, 1970, p. 3112.

[37]*Congressional Record,* April 15, 1970, p. 11885.

[38]The methodology on which that claim was based has been the subject of considerable criticism, and the findings were contradicted by those of the more carefully designed Seattle-Denver income maintenance experiments. See Philip K. Robins and Richard W. West, "Labor-Supply Response Over Time," *Journal of Human Resources,* 15 (1980), p. 541.

[39]Finance Committee, Senate, Hearings on H.R. 16311, 1970, p. 914.

[40]William A. Morrill, "Introduction," Symposium on Graduated Work Incentives, *Journal of Human Resources,* 9 (1974), p. 156.

[41]Albert Rees and Harold W. Watts, "An Overview of the Labor Supply Results," in Joseph A. Pechman and P. Michael Timpane, eds., *Work Incentives and Income Guarantees,* p. 87.

[42]Michael C. Barth, Larry L. Orr, and John L. Palmer, "Policy Implications: A Positive View," in Ibid., p. 207.

[43]David H. Greenberg and Philip K. Robins, "The Changing Role of Social Experiments in Policy Analysis," *Journal of Policy Analysis and Management,* 5 (1986), p. 345.

[44]Barth et al., "Policy Implications," p. 207.

[45]See Ellen Tenenbaum and Aaron Wildavsky, *The Politics of Mistrust,* Sage Publications, 1981.

[46]Vincent J. Burke and Vee Burke, *Nixon's Good Deed: Welfare Reform,* New York: Columbia University Press, 1974, p. 132.

Discussion

Hugh Heclo*

The Coyle/Wildavsky paper asks why an effort to radically reform the welfare system should have failed. Their answer is derived from a tripartite view of American political culture. Essentially, the egalitarians were unwilling to accept half a loaf.

First point. It is not clear why it is so necessary to rev up cultural theory and overheat the word processor to explain the fate of welfare reform in 1969-70. We can expect that it will be difficult to persuade Americans (elite or mass) about the merits of a plan for transferring some of their money to an easily despised fraction of the population.

Second point. Taking the premises of the analysis, the conclusion of the paper is not only self-evident but preordained. Given that America is composed of three cultures; given that radical change is defined as a major alteration in relative power; it follows that radical reform will require the consent of the three cultural blocs. Q.E.D. Of course this country has shown it is possible to obtain radical change by playing two against one. But that is not called reform. It is called civil war. Working with the Coyle/Wildavsky formulation, we can only wonder why historians have spent so much time debating the causes of the Civil War and missed the fact that it was the egalitarians (Abolitionists) and individualists (Northern capitalists) ganging up on the hierarchs (guess who).

Third point. While I sincerely believe there are substantial insights to be derived from the recent movement to apply cultural concepts to American politics, I must say that this paper tells us little about the

*University Professor, George Mason University; formerly Professor of Government, Harvard University.

historical and political context of the negative income tax experiments as such, the focus of this conference. We are told that at times the results of the experiments were used as political ammunition and that those backing the experiments may have been a bit naive about the relationship of social science findings to policymaking. No news here. Working within the confines of the paper, the relevant—and unaddressed—question is this: What is the relation between this idea of three cultures on the one hand and the creation and operation of the income maintenance experiments on the other? If this were older history and none of the participants in question were alive to argue back, it might be an easier question to answer. Would all of the egalitarians pushing negative income tax experiments in the 1960s please stand up? Since this whole effort at performing deliberate social experiments on Americans was a pretty radical departure for the federal government, did not the individualists and hierarchs in the audience have to sign on too? When it comes down to it, who among us thinks that his or her views can be jammed into one of these three pigeonholes?

I believe there are important issues lurking in the larger context of these experiments. Unfortunately they are not revealed in this paper. The phenomenon of social experimentation was itself a sociopolitical experiment. What was happening here—was something worth intellectual attention going on beyond the particular experimental findings? Forget the substantive results for a moment. What did this phenomenon *mean?*

Thankfully, I have not been given the job of trying to write such a paper. In the remaining space allotted to me, and drawing as best I can on several other papers, let me offer one possible sketch.

The negative income tax experiments represented the triumph of what Richard Elmore in his paper terms an analytic subgovernment. No politician in the White House, no congressman, no interest group as conventionally defined and no lobby of rank and file citizens was pressing for a multi-million dollar system of negative income tax experiments. Their creation was the work of a more or less autonomous economics profession and a particular school within economics at that. One part of the story is how their intellectual constructs came to prevail in this postwar period. A more obscure but no less important part of the story is how their closest disciplinary competitors for thinking about the poverty issue—social work/sociology—collapsed from within. It is a story hinted at in the Coyle/Wildavsky paper's mention of social workers' reaction to the proposed Family Assistance Plan. A discipline filled with such loathing for its own tradition was simply no match for the economists.

That the income maintenance experiments could happen in this way tells us, I think, something even more important. It tells us how far we have come from our original vision in this country about the role of social science. One hundred years ago, the founders of America's modern system of social investigation (Spencer Baird, Otis Mason, Ainsworth Rand Spofford, John Eaton, Francis Amasa Walker, John Wesley Powell, Lester Ward and Carroll Wright, to name a few) saw social inquiry as a new kind of instrument for linking state and civil society. Social science was seen to have a civic purpose. By contrast, the massive machine of negative income tax experimentation can be seen as an indicator of just how far apart have been drifting the separate realms of politics, social science, and the understandings of ordinary citizens.

Existing in this kind of splendid isolation, the negative income tax experiments represented a centralizing, reductionist impulse—a search for the one right answer—that comes naturally to a single disciplinary view of the world. Only from this mindset could it make sense to spend millions of dollars on four experiments to see if people worked less in response to income guarantees and next to nothing to find out what they did with any lessened time on the job. So much for economists' supposed preoccupation with scarce resources.

As we all know, the negative income tax mentality (the wit of NIT?) has gone into remission and no talk of income guarantees for the poor is to be heard in the land (the non-poor are another matter). What has been left in its wake? These social experiments themselves became part of the new historical and political context. On the one hand the whole episode probably contributed to the no less monofocal view of those reacting against income guarantees, against "incentives" for proper behavior rather than punishments for bad, against the dependency-creating effects of poverty programs, and so on. As for the negative income tax experimenters, at least it can be said that they were trying to find answers to questions about which they were not sure. Their conservative successors in the monofocal sweepstakes of American politics are more interested in bringing an indictment based on the way in which they *know* the world works.

On the other hand, and of longer lasting importance, the negative income tax experiments broke ground for a succeeding generation of social experimentation. It is a generation not only of more refined techniques but also, as Barbara Blum's paper reminds us, of more sensible connections to existing political and administrative structures. The era of the single, dramatic, Washington-centered experiment is gone, gone as quickly as it arrived. In its place is the more familiar pluralism of social inquiry involving state and local governments, foundations, more disciplines. I suppose we could take the author of the concluding paper, Dick Nathan, as a representative character in this evolution. His migration from the old New Federalism of revenue-sharing to the central income guarantees of the Family Assistance Plan, and hence to the fine-grained Manpower Demonstration Research Corporation experiments with services and multi-program approaches, tracks rather well with the central tendencies of our time.

If we are willing to pause for a moment and look past the experimental findings, the controversies about workfare, the budgetary pressures on new research funds . . . if we are willing to be so untopical as to even look past the hot ideas for any next round of welfare reform, what we will see in the last 20 years is a society more busily engaged than ever in seriously trying to know itself.

Discussion

*Lawrence M. Mead**

Authors Dennis J. Coyle and Aaron Wildavsky argue that welfare reform did not pass because it failed to satisfy the conflicting cultures in American politics. Major reform in the United States, they contend, must usually appeal to all three of these cultures. Their ideals are respectively hierarchy (a concern for social order and due place), egalitarianism (equality of condition), and individualism (equal opportunity).

These categories have the appeal that they exhaust the major ideals to be found in political theory. Behind them lie the great names of Burke, Rousseau, and Locke. But I would question whether they are equally rooted in American politics. Where are the American hierarchs? I thought that aristocratic visions of society went out of American politics no later than 1800, when Jefferson defeated the Federalists. Since then, even the right wing in American politics has had to appeal to the people.[1]

It is true that American government involves a degree of hierarchy and authority. The New Deal still sets the frame of American politics, and New Deal politics was highly organized. Large-scale political structures, both parties and interest groups, mediated the demands of the people to government, and New Deal policy used public bureaucracy on a new scale to serve the people.[2] The social vision, however, was a Lockian one, and not a Tory one as the term hierarchy might imply. Government was still the servant, not the master, of the people. Its aim was not to enforce social inequality but to ameliorate it, albeit by steps well short of socialism.

And as the authors suggest, even governmental hierarchy has since come into question. After 1960, an egalitarian politics of protest and

*Associate Professor of Politics, New York University.

single-interest groups undercut the traditional parties. Community Action and other innovative grant programs inaugurated a new, anti-bureaucratic style of federal public administration.[3] These trends were answered in turn, not by a reaffirmation of hierarchy, but by a recrudescence of free-market conservatism. Today, it would seem, the egalitarian and individualist persuasions—those that denigrate authority—dominate the political culture.

Coyle and Wildavsky suggest that welfare reform failed because the three cultures refused to compromise. The egalitarians demanded a degree of income guarantee unacceptable to the individualists, while the hierarchs wanted to enforce values, especially a work requirement, that were unacceptable to either of the others. In part, I agree. This analysis is certainly an improvement over the view, emphasized in other accounts,[4] that reform failed largely because of conventional "New Deal" disagreements between left and right over the scale of government.

But this view fails to explain the most startling thing about welfare politics—the fact that the public is nowhere near as divided as the elites. The public lines up unequivocally with what Coyle and Wildavsky call the hierarchical position. It wants to guarantee the poor jobs rather than income, and this preference extends even to minorities and the poor themselves. The polls that the authors cite positively radiate the desire to *enforce* at least minimal norms through public authority, the animus that is so lacking in the other two cultures.

Some Congressmen spoke for this position in the Family Assistance Plan debates. I call them the moderates or civic conservatives. Their leaders were Martha Griffiths and Russell Long. It could even be said that they defined the consensus toward which the debate progressed. Over time, the welfare plans in Congress relied more on requirements, less on incentives, as the mechanism for promoting work by the employable. But the moderates were outnumbered, and eventually disagreements among all the groups killed reform. The same disputes, along with a greater fear of costs, killed Jimmy Carter's welfare plan, the Program for Better Jobs and Income.[5]

There seems to be a division between what I would call social opinion and political opinion. Unpolitical Americans are the ones who speak for hierarchical values, who want to use government, not just to help people, but to enforce the civilities essential to American life, one of which is work for the able-bodied. Most active politicians, however, want government to serve the values of freedom and opportunity rather than order. Liberals locate the main barriers to freedom in private society, while conservatives find them in government itself. Liberals therefore want a larger government, to protect people from the economy, while conservatives want a smaller one, to give the economy full sway.

The public seems to be little interested in these disputes. A study by Verba and Orren shows sharp polarization between liberal and conservative groups over whether "the system" or the poor themselves are responsible for poverty.[6] But these are the views of elites, bound up in the New Deal struggle over the scale of government. According to other studies, unpolitical Americans reject this polarity. They hold *both* government and the individual responsible for social problems. They are much less interested in changing how much is done for the poor than in enforcing decent behavior on those who are helped.[7]

Why were the politicians not as unified as the public about welfare? In the authors' terms, why is the hierarchical persuasion much the strongest among ordinary citizens, while individualist and egalitarian visions that deny the need for public authority flourish among elites? Concretely, *why do politicians not do in welfare what the public wants,* which is to guarantee the needy a sustenance but make the employable work for it?

That is the great mystery in welfare politics. Merely to label the persuasions as different cultures does not account for it. My own view is that it must go back to the founding of the country. The framers of the Constitution, alone of founding elites, construed their task as the limitation rather than the enhancement of national public authority. They presided over a healthy society in which reform at the hands of government never would be as necessary as in the modernizing societies of Europe (or now of the Third World). They construed government as the product of society, rather than society as the creation of an enlightened government. Ever since, the main political dispute has been how, not whether, to subject public authority to the people. That has made it tougher to *use* government for the tasks of social reconstruction that exist even in America, of which the most daunting today is integrating the welfare class.

In welfare, reform failed partly for pluralist reasons. The authors mention that specific groups—for example, social workers, Southern politicians, the National Welfare Rights Organization—would have lost from it and thus opposed it. Welfare is certainly a subject on which consensus is notoriously elusive. The fundamental reason for defeat, however, was that the reforming elites were out of step with public opinion. They would not or could not reform welfare in the way ordinary Americans wanted. They would not enforce social obligations like work in return for benefits.

Social order as a concern entered prominently into welfare politics. Welfare is a fundamental disorder in American life. Long-term dependency is offensive to the American social vision of a nation of equivalent citizens. It is also a cause of other disorders such as crime, drug addiction, and the decline of the schools. Thus, the public has strong views about it and demands that any reform reduce disorder.

In welfare, the central issue is not who should be subsidized or who should win or lose, but how to elevate the seriously dependent so that they can even play the political game like other people. Welfare politics is abnormal. The question is how to create the community that is assumed in ordinary politics. It is how to make the poor self-reliant enough so that the stakes of politics are no longer critical.

Abnormal politics is much more distressing to the public than ordinary politics. It raises basic issues of personal and social identity that ordinarily never come up. Pressures for change arise, not from economic claims, but from social dysfunction. Claims arise, not from competent economic interests, but, so to speak, from the disassembly of the personality. The long-term dependent do not have their lives "together" enough to be the kind of individuals imagined in either the individualist or the egalitarian vision of society. Whoever is to blame, they threaten social order at a much more fundamental level than anything about the tax system, for example.

Questions of social order expose the limitations of federal governance. Whether we speak of a New Deal division or competing individualist and egalitarian visions, all the dominant tendencies in Washington seek to assume what in social policy must be created—a nation of competent citizens. Would that they listened more to the civic attitudes articulated by public opinion. The problem in social policy is to make government as civic-minded as society.

There is an additional problem too. The serious claims in modern politics are about dependency, and yet elites have not learned to talk about them rigorously. We have a language for discussing claims to political freedoms. It is the language of democracy and civil liberties, the language of the seventeenth and eighteenth centuries, of what one might call middle-class politics. But—except in South Africa—these issues are passé. We also have a language for discussing claims to economic protection. It is the language of socialism and collectivism, the language of the nineteenth and twentieth centuries, of what one may call working-class politics. But, despite the ambitions of the Reagan administration, the welfare state is established and its boundaries are unlikely to change much anywhere in the West.

We do not have a language to discuss the claims that arise from the appearance of an underclass in Western societies. What does one do about the social dysfunction that remains even *after* a society has carried out the reforms specified in middle-class and working-class politics? Who is to blame for serious dependency, and what is to be done about it? Disputes revolve around claims to determinism, not claims to freedom in the earlier sense. Those who speak for the poor assert that they are dominated by their environment and not responsible for dysfunction, while their opponents deny it. The earlier theories all assumed that even the downtrodden were accountable for decent personal behavior, while in dependency politics, that assumption is itself the main issue.

Perhaps the final reason that welfare reform did not succeed is that welfare raised embarrassing issues of personal adequacy that politicians *hate* to talk about. The moral of past welfare reform is that we have to talk about them. We need serious arguments, based on research, that go beyond rhetoric on the serious behavioral questions in welfare. Who is and who is not responsible for personal functioning, and for what specific competences? What is the potential for human nature to achieve civility?—the very question that conventional political and economic theory never asks.[8]

Such languages do not eliminate disagreement, but they clarify views and, over time, narrow the differences. Consensus can then be embodied in policy. Recent discussions of the welfare problem have been newly open about discussing the behavioral problems, and that is a step on the way. The goal is a political theory and a constitutional doctrine about permissible degrees of dependency, such as we already have for civil liberties and economic regulation. Only on this basis could government set a standard for behavior on welfare and enforce it, as the public wants.

---

[1]Louis Hartz, "The Whig Tradition in America and Europe," *American Political Science Review*, 46 (December 1952), pp. 989-1002.

[2]Richard Hofstadter, *The Age of Reform: From Bryan to F.D.R.* (New York: Knopf, 1955), ch. 7; Samuel H. Beer, "Liberalism and the National Idea," *The Public Interest*, 5 (Fall 1966), pp. 70-82.

[3]Samuel H. Beer, "In Search of a New Public Philosophy," in Anthony King, ed., *The New American Political System* (Washington, D.C.: American Enterprise Institute, 1978), ch. 1.

[4]For example, see Daniel P. Moynihan, *The Politics of a Guaranteed Income: The Nixon Administration and the Family Assistance Plan* (New York: Random House, 1973).

[5]Lawrence M. Mead, *Beyond Entitlement: The Social Obligations of Citizenship* (New York: Free Press, 1986), ch. 5.

[6]Sidney Verba and Gary R. Orren, *Equality in America: The View From the Top* (Cambridge, Mass.: Harvard University Press, 1985), p. 74.

[7]For a summary, see Mead, *Beyond Entitlement*, pp. 233-40.

[8]A pioneering inquiry along these lines is James Q. Wilson and Richard J. Herrnstein, *Crime and Human Nature* (New York: Simon and Schuster, 1985).

A Sociologist's View of the Income Maintenance Experiments

*Lee Rainwater**

Given how much money was spent on the negative income tax experiments, what can be learned about social, as opposed to economic, behavior seems remarkably skimpy. Nevertheless, something interesting can be learned from a consideration of why the experiments tell us so little about the people who were their subjects. I may have a vested interest in making this assertion, since when the experiments were being planned I argued that such would be the result. That is, we would not be able to understand in human terms what had happened because of the narrow way in which the data collection was to be done — and now I can say, "I told you so."

One might bring to bear a sociological view on the negative income tax experiments in one of two ways. The first is to ask what can be learned from the experiments that is useful in the development of substantive or theoretical sociological knowledge. The other is to ask what sociology has to contribute to policy lessons. Most of my comments are concerned with the latter, but from time to time I will take account of the reverse flow from "applied" to "basic" knowledge and puzzle over why it is so frustrating to try to learn something about social behavior from these experiments.

My remarks will offer ample evidence of a distinction often drawn between economics and sociology. Economists are interested in the choices people make; sociologists are interested in the fact that people have no choice. Our model is that constraints by institutions, culture, interaction dynamics and personality determine and overdetermine behavior in ways that often leave little room for rational choice. One of the important lessons for sociology as a field from the experiments has

*Professor of Sociology, Harvard University.

to do with understanding that choice still plays a role.[1]

I want to deal with three aspects of what I see as the failure of the experiments. The first concerns a failure of perspective in the initial conceptions of the negative income tax as a policy initiative: the experiments were done because of the widespread belief among policy experts that the negative income tax could bring major benefits to society. (I do not mean that only the other guys thought so; I supported the initiative also.) To properly see this failure we need to broaden our policy thinking in ways that have proved particularly difficult for American policy specialists.

The second lesson to be drawn has to do with a failure of method. Why, after the great expenditure of talent, time and money, have we learned so little? What are the lessons for social science methodology that we should take from a consideration of the quality and quantity of the experimental findings?

Finally, there is the failure of policy interpretation from the experiments. We should and could have learned much that we did not. There are reasonably solid findings in the area of labor supply and shaky but plausible findings in other areas, summarized in the three papers prepared for this conference by Burtless, Hanushek and Cain. The established, conventional wisdom concerning the experimental findings is that any negative income tax would be a very expensive and a very pernicious program. From a broader perspective, what other kinds of policy implications might be drawn?

Policy Paradigms That Informed the Experiments

It is easy to lose sight of the context of social and political concerns that led to an interest in the negative income tax as a policy innovation, but this context should inform our assessment of the lessons to be drawn. Remember that the idea of a negative income tax was put forward by a large group of economists of highly varied political persuasions as a centerpiece for a sensible war on poverty. It was also a distinctly American idea. Neither at the time nor since have other countries shown enthusiasm for a negative income tax.

The belief that something had to be done about poverty was clearly related to the revolution in race relations going on in the country. The growing realization of the plight of blacks outside the South was particularly important. In the early sixties the warnings of the Triple Revolutionists concerning automation and accompanying unemployment had captured a good deal of attention among intellectuals. Their proposals for a guaranteed annual income brought that idea to the fore. Though their arguments about the inevitability of rising unemployment were eventually defeated in the debate that followed (see Robert Lekachman's contemporary summary), the notion that job creation strategies for combating poverty were either too expensive or doomed to failure had taken hold by the end of the decade.[2] It seemed obvious that the lower class, and particularly blacks as the most disadvantaged victims of this process, would be hard to help with job creation programs.

Therefore, a prosperous society could meet its obligations to the poor more cheaply with transfer programs than with programs that required changes in the way labor market institutions operated. One heard that it was cheaper to give money to the poor than to make jobs for them. The negative income tax won out over the credit income tax, its main income maintenance competitor, with the argument that money could be sharply targeted on the poor, and thus the tax burden kept lower.

By the mid-seventies the national mood had changed, and it no longer seemed urgent to do something to change in any basic way the propensity of American institutions to generate a rather steady rate of relative poverty — probably around one-sixth to one-fifth of families since the 1950s.[3] During the civil rights revolution the establishment mood had combined sympathy and not a little fear. Now the fear was gone and a nasty mood of "We've given them too much already" took its place. In this context the experimental findings of work and marital stability disincentives provided welcome support, but it would be a mistake to think that they caused the turning away from an incipient national commitment to increase the income of the poor.

As a way of appreciating the distinctiveness of these policy paradigms, let me return to the earlier observation that the guaranteed annual income never excited much interest outside the United States. European welfare states have generally adopted programs earlier than we, and they spend more of their national income on the programs we have in common. Why did the negative income tax not excite interest in the same way that many American social policy ideas have—for example, community action?[4]

The dominant model that has guided the development of the European welfare state has been a social security model. It has assumed several things about the society and the economy. The most important was that full employment would be a central goal of the state, and that it would exist most of the time. For the most part it was expected that the income distribution generated in the economy was a reasonable one, and that the purpose of social protection programs was to maintain a reasonable approximation of a family's usual income when an untoward event reduced the capacity of breadwinners to work—sickness, injury, old age. Family allowances existed to compensate families who chose to rear children and thus made more equitable the distribution of income between parents and nonparents. But family allowances were not expensive programs. The allowances generally have not increased as fast as other benefits; in most countries they did not keep pace with the growth in real income.

The assumption that social programs replace adequate work income, rather than substitute for it, not only requires near full employment, but also a very small low-wage sector. In Sweden, which has no minimum wage, almost no one who is employed full-time will be poor because the lowest wage is above the poverty line. In most other advanced welfare states in Europe the lower half of the income distribution has much less variance than in the United States (France may be an exception). The combination of a small low-wage sector and near full employment until recently meant that there were few pre-transfer poor working-age families in Europe.[5]

We have fairly precise evidence concerning the proportion of poor families for which a negative income tax would have been designed in Sweden, Britain and Germany compared to the United States around 1970.[6] Among families with a head age 25 to 54 years, the pre-transfer poor ranged from 6 percent to 9 percent in the European countries and amounted to 14 percent in the United States. The post-transfer poor ranged from 1 percent to 3 percent in Europe and 11 percent in the United States. Germany and Sweden moved three-quarters or more of their pre-transfer poor out of poverty with transfers, Britain 58 percent, and the United States only 22 percent. Few Europeans required public assistance to move them from poverty, and thus very few would have been candidates for a negative income tax.

These differences between Europe and the United States persist to the 1980s. Data from the Luxembourg Income Study (using a different poverty line and equivalence scale than our 1970 study) show that European post-transfer poor families with heads age 25 to 54 years ranged from 5 to 7 percent, compared to 15 per cent in the United States.[7] Poverty among children is found to range from 5 percent to 11 percent in the European countries compared to a 24 percent U. S. rate.[8]

From the mid-1970s this standard assumption of European welfare strategy — that social programs could rely on economic growth, full employment, and a compressed wage distribution from the low end to the middle — has increasingly been called into question by economic events. One might have thought an interest in guaranteed income strategies might have arisen on the part of mainstream policymakers and intellectuals. But it has not. Instead, policy responses have continued to emphasize the employment link through extended unemployment benefits, emphasis on subsidies to industries to maintain employment, and extensive retraining programs. Sweden provides the most dramatic example — expenditures on employment-related social programs increased from 0.5 percent of GNP in 1965 to 2.6 percent in 1983. (Benefits from industrial subsidy programs should be added to this, but data on numbers of beneficiaries and level of benefits from these programs are not available.) During the 1970s Sweden also expanded considerably its housing allowance program, which works like a negative income tax, but by 1983 it still cost less than 0.4 percent of GNP. Child allowances and advanced maintenance payments amounted to 1 percent of GNP.

Thus, about 4 percent of Swedish GNP is directed to the kinds of families who might be candidates for a negative income tax, but the mechanisms are very different, reflecting the social citizenship orientation of the Swedish welfare regime in contrast to the public assistance orientation of the American regime. In Sweden, child allowances go to all families; employment programs are for all those without adequate jobs; industrial subsidy programs slow down the pace of rationalization; housing allowances have high break-even levels and are administered in such a way that no stigma is attached to their receipt. (In any case the average beneficiary is not heavily dependent on them.) These program aspects are a reflection of a broader set of policy choices which together constitute a particular country's welfare regime.

In short, the first lesson to be drawn is a political-sociological one. It is essential to see any policy research in the context of actual or potential policies involved, and the chosen policies, in turn, in the context of the overall welfare regimes of which they are a part. The choices nations make can only be seen if we look at national policies in a comparative context.

Methodological Disappointments

The point of research is to increase understanding and, if possible, control. The experimental method in the social sciences, at least when done without a rich context of substantive knowledge on its subject, ends up long on control and much too short on understanding. The experiment is a black box — we know what goes in, and we know what comes out, but we don't know what went on in between, nor do we know how what happened happened. If the control the experiment provides does not also yield understanding, then we probably will do a bad job of prediction.

I think the principal shortcoming of the policy studies called negative income tax experiments was to conceive of them as only experiments. They were necessarily much more, and should have been even more than they were. As the researchers quickly discovered, they were forced to invent a whole system of administration and control — the experiment was also a program, a mini-program. All of the problems of implementation were present.

If the experiments had been conceived as a study in which was imbedded an experiment, we would know more. Instead the black box was preferred, and therefore we cannot answer interesting questions about the findings. Why did the marriages that broke up break up? What was the nature of those marriages that broke up, among both controls and experimentals? Were the people who reduced their work effort conscious of doing so? Did they think they did so because of the experimental payments? What did they think their own purposes were? Why did they not take all the money they could get by combining working and payments? What was their reasoning? More broadly, what did they think was going on — how did they construe the experiments? Early on, how did they expect to be affected, and how did their conceptions change over time?

In short, what is missing from the experiments is description, the most primitive level of science. The more ambitious goals of science, we all learn in introductory texts, presume the less ambitious. Without description we have hard findings, but brittle ones. More description might have yielded tougher findings. We would know much more if more resources had been devoted to chronicling the operation of the experiment (administration, relations of staff to subjects, etc.) and to learning more about the life situations of both experimentals and controls over the time of the experiment.

To do this, however, would have challenged much that is central to the ideology of contemporary social science. Much of the data that were needed would have been qualitative, narrative, processual. But quantitative methods have captured the imagination of both economics and sociology. We want so badly to be respected as sciences, and we have fixed on the elegant manipulation of numbers as the way to gain that respect.

The economists who ran the experiments might not have been willing to sponsor methods that looked "soft" and "subjective." Given the centrality of revealed preferences to economic methodology, they might have been little interested in rich contextual analysis of what was going on in the heads of the families being studied. But they had little chance to make the choice. For a couple of decades before the experiments, social psychology, for example, had pursued the quantitative will o' the wisp into a dead end, so that it had very little to offer in the way of useful methods to measure some of the noneconomic matters of interest. Sociology has forgone the opportunity to develop flexible methodologies of open-ended interviewing and coding in favor of closed-ended questionnaire approaches that seemed to offer more scientific control (and were cheaper, too).

The result is that a great many questions were asked, many answers given, recorded, tabled, but we never get to know the people and their lives.

The Failure of Policy Interpretation

The narrowness of the policy conclusions drawn from the experiments follows from the two kinds of problems sketched above. Because our knowledge from the experiments is shallow, and because the experiments are conceived in the context of the American welfare regime, with its implicit assumption of high unemployment rates, inactive manpower policies and a means-tested bias in transfer programs, the ''logical'' conclusion that is drawn from the findings is that the negative income tax is not a useful policy innovation. After all, we do not want to encourage the breakup of families, and we do not want to discourage people from working, and we do not want to spend a great deal of money on the poor.

Because we do not know anything about the dynamics of the marital disruptions associated with the negative income tax, even if one accepts the finding it is hard to draw other than simple-minded policy implications. Suppose all the excess of disruptions involved battered wives, or wives of alcoholic men. Would the policy implication still be that the negative income tax has an unconstructive effect? Do we believe that the extra money caused people to be self-destructive in the sense that they ended unions in which they would have been better off remaining? What happened to consumer sovereignty? If we want to give couples the best chance to stay together and if more income will make things better, we have only to choose the more generous plan.

To decide what the policy implications — as opposed to cost implications, the two should not be confused — of reduced worktime are, we need to know more about when people reduced effort, and what they did with the time. If the unemployed were a little slower in finding a job, they must have done some other citizen who wanted a job a favor. Given our unemployment rates, recipients taking their time finding a job because they receive negative income tax payments isn't likely to create a labor shortage. The marginal reductions of working hours increase the cost of a negative income tax but they do not seem large enough to constitute a major change in low-income workers' attachment to the labor force.

Given the ambiguity of the findings, whether the negative income tax was a good idea for this country reduces mainly to a question of cost — or of choice. From a comparative perspective on social protection even the most generous plan described by Gary Burtless would be relatively cheap — just 1.5 percent of GNP. Almost all European countries spend far more than we on social protection programs, yet their societies have not fallen apart and, within their economic realities, their economies are in reasonably good shape. Most have seen the proportion of GNP devoted to such programs increase over the past decade rather than decrease as ours has done. If for reasons of our particular history we were not likely to go the European route of heavy investment in employment programs, modest child allowances for all families, and even more modest housing allowances, then a negative income tax could have been considered a sensible, pro-family addition to the nation's social protection programs.

---

[1]It is also encouraging to see an economist like Thomas Schelling begin to deal with issues of "nonrational" personality constraints; I hope that other economists will expand this kind of interest to more social constraints. Cf. Thomas C. Schelling, "Self-Command in Practice, in Policy, and in Theory of Rational Choice," *American Economic Review*, 74, (May 1984), pp. 1-11.

[2]Robert Lekachman, *The Age of Keynes*, McGraw-Hill Book Company, 1966, pp. 226-245.

[3]The one-sixth figure comes from my unpublished study using survey data starting in 1953. The one-fifth figure comes from the census data analysis by Christine Ross, Sheldon Danziger and Eugene Smolensky, "The Level and Trend of Poverty, 1939-1979," IRP Discussion Papers: DP#790-85, December 1985.

[4]See Peter Maris, *Community Planning and Conceptions of Change*, Routledge and Kegan Paul, 1982.

[5]See Michael Bruno and Jeffrey D. Sachs, *Economics of Worldwide Stagflation*, Harvard University Press, 1985, for a discussion of some of the consequences of high wages under recent economic conditions.

[6]Lee Rainwater, Martin Rein and Joseph Schwartz, *Income Packaging in The Welfare State*, Oxford University Press, 1986, and unpublished German data.

[7]Peter Hedstrom and Stein Ringen, "Age and Income in Contemporary Society: A Comparative Study," L.I.S.-C.E.P.S. Working Paper Series # 4, CEPS, Walferdange, Luxembourg.

[8]Timothy Smeeding, Barbara B. Torrey and Martin Rein, "The Economic Status of the Young and the Old in Six Countries," paper presented at annual meeting of the American Association for the Advancement of Science, May 1986.

Discussion

*Charles Murray**

Lee Rainwater's paper makes some points that I want to second without elaboration. Program assessments—and the negative income tax experiments were a demonstration program writ large—have been persistently myopic in all the ways Rainwater points out. They have ignored the policy context within which they must fit. They have shown a remarkable indifference to trying to penetrate the black box of causation. They have been downright hostile toward qualitative data.

I do wish to elaborate on Rainwater's last and quite important observation, that the negative income tax findings do not drive an answer to the question, "Is a negative income tax a good idea as national policy?" Rainwater generally refers to the ways in which its advocates on the left might be able to live with reduced labor supply and increased marital disruptions. But the same point could be made of its advocates on the right. A Milton Friedman, one of the earliest proponents of a negative income tax, might well read the evaluations of the experiments and nonetheless continue to support the concept—Friedman always expected that a negative income tax would produce work disincentives, and the results of the experiment do not negate (or say much about) the merits he originally saw in replacing a Rube Goldberg welfare system with a tidier one. Has Friedman changed his mind because of the experimental results? I do not know, but nothing in the findings would demand that he do so.

Let me add to Rainwater's list three other reasons why we still have little idea whether a negative income tax is good policy or, for that matter, whether a wide range of other antipoverty devices are good policy.

*Bradley Fellow, Manhattan Institute for Policy Research.

First is the problem of deciding upon a baseline. By definition, a negative income tax must have one. For the negative income tax experiments as in most other means-tested programs, the poverty line served that function. Let me suggest that the poverty index (along with any conceivable refinement of it) is so inadequate—so completely meaningless—that it obscures both the interpretation of the negative income tax results and any inferences about policy to be drawn from them. When I say "meaningless" I am not referring to problems of valuing noncash benefits, or to problems of imputing unreported income, or to any other marginal technical problem. I mean "meaningless." To know that someone is below the poverty line in this country is to know extremely little about that person except for the most general inference of low income.

The dominant source of the meaninglessness of the poverty index is the difference between living a low-income life in the inner city and in a small town. The average poverty index for a family of four in 1985 was roughly $11,000. Any of you in this audience with a spouse and two children, given that amount of money and told to live on it in a small town in Missouri or Colorado, could make a decent life for yourself and your family. You could get a decent place to live—small and shabby perhaps, but one that could be kept clean, warm, and dry. You could eat nutritious food, send your children to a pretty good school, live in a neighborhood with stable families and employed fathers and well-brought-up children. You could be safe from criminals. If you got the same $11,000 and had to go live in the South Bronx, you could not make a decent life for yourself. You could not do it, I submit, if your income was twice or three times the poverty line.

This source of error in the poverty index is not going to be finessed by more sophisticated cost-of-living discriminations, because the source of the problem is not the difference in the cost of hamburger in the South Bronx and in a small town in Missouri. It is the difference between a civilized, functioning community and one that is lawless and foundering. The poverty index cannot be reconstructed to cope with this. There is no negative income tax that can establish a baseline income that will enable everyone to have a decent standard of living. Persisting in attempts to correct any baseline index will probably only make matters worse—the most likely result being the introduction of large dollar differences in the allowances for urban and nonurban areas, in effect bribing people to stay in places like the South Bronx without making it possible for them to control their environment.

Another reason why we have no basis for voting yea or nay on a national negative income tax on the evidence of the experiments is that no one (to my knowledge) has used much imagination in anticipating the policy choices that would arise if the experiment were implemented nationwide. Let me give just one example of many. If a national negative income tax were put in place, we can predict (on the basis of what we do know about the nature of the underclass) that at the end of the first month, very large numbers of people will be without money for food. Large numbers of people will not have paid their rents. For good reasons or bad, large numbers of persons who have enough money to live a decent existence will be living the same existence they live now. What shall we do? Install Son of Food Stamps? Create a Special Rental Assistance Program for persons unable to manage their money? The merits of a negative income tax that supplants the current system are one thing; the merits of a negative income tax that supplements the current system are quite another. Professor Rainwater has lamented that we failed to consider policy in the context of the welfare state in which the policy must be implemented; I am lamenting that we failed to consider it in the context of the welfare population who would be the major intended beneficiaries of the program.

Finally, let me suggest that the assessment of the negative income tax, in common with the assessment of every other social program, wears blinders when it comes to selecting dependent variables for measuring success. If the negative income tax had produced very small changes in labor supply and if it had not shown other deleterious effects such as marital disruption, it surely would have been interpreted as a great success. It would have accomplished the Great Good of recent social policy, bringing people above the poverty line. But just because income is one of the few outcomes we can measure reliably, it does not follow that effects on income can be segregated from all the other good things that we would like people to have in their lives. On the contrary, it seems more reasonable to assume going in (it is always the safest assumption) that economic and noneconomic rewards interact, rebound off each other, and that gains in one area may well come at the expense of other equally important objectives.

We wallow in great ignorance on such subjects. For example, when curmudgeonly conservatives like me raise the possibility that it is fundamentally important that a person *earn* his own income—that earning one's own keep and pulling one's own weight are the source of deep-seated rewards and satisfactions, especially for persons who do not enjoy the limelight in any other respect—the hypothesis is too often presumed to be adequately tested by running out and administering an eight-item self-esteem scale to recipients of income transfers and then, when they seem to score as high as anyone else, saying, "Well, we dealt with that."

I would argue that we can do better, but to do so we will have to develop a richer interchange of ideas, knowledge, and methods across disciplines. I am thinking of economists and sociologists and political scientists, but most especially of a group not represented here, psychologists. Fascinating literatures are accumulating on such topics as locus of control, on the role of "competence" and "self-determination," and on the complicated relationship between intrinsic and extrinsic rewards. These bodies of research can be pulled into our ken. They inform the things we want to accomplish with programs such as the negative income tax. In some cases, they tell us about ways to provide benefits that can be more fruitful than the ones we might intuitively choose. In other cases, they can at least permit a knowledgeable dialogue about the trade-offs between income and the other aspects of quality of life that we have not systematically considered in the past.

So we have a considerable way to go, in my view, before we know what to make of the results of the negative income tax experiments. We need a new baseline for defining where "negative" begins in "negative income tax." We need to think in more detail about how a national negative income tax would differ from an experimental one. And we need to employ broader and more sophisticated dependent variables in assessing what we are accomplishing. Lacking those, we are not much better able now than we were 15 years ago to tell whether the negative income tax would be good policy.

A Political Scientist's View of the Income Maintenance Experiments

*Richard F. Elmore**

The negative income tax experiments were designed to influence political debate on income support policy in at least two ways. One was methodological — to focus the policy debate on a few key empirical questions and produce more definitive evidence than would have been available through nonexperimental methods. Another was political — to legitimize the idea of a universal cash transfer program, scaled to the recipient's income, as an alternative to the patchwork collection of in-kind and categorical assistance programs that had grown up since the New Deal.

Methodological Issues

Social experiments are thought to hold certain advantages over nonexperimental policy research. They frame policy questions in more precise terms. They permit more precise statistical tests of effects. And they introduce a dimension of greater empirical discipline to discussions of policy issues.

These advantages are, however, purchased at some cost. The very rigor of social experiments limits the utility of their results. First, experimental treatments are, of necessity, packages of discrete elements — in the negative income tax experiments, various combinations of guarantee levels and marginal tax rates. Inferring the effects of treatments not represented in the experiment from experimental results

*Professor of Education, Michigan State University and Senior Research Fellow, Center for Policy Research in Education.

requires extrapolation between or beyond treatments. This extrapolation is no more rigorous than most nonexperimental research. Second, treatments are, of necessity, implemented in specific sites, which means that experimental results have to be combined with nonexperimental survey data to estimate the effects for the population as a whole and to estimate the cost of generalizing the treatment. Combining experimental and nonexperimental results reintroduces most of the methodological tangles that experiments were designed to eliminate. Finally, experiments take place at particular times, against a background of particular policies and particular economic, political, and social conditions that influence the responses of subjects. These background factors cannot be controlled experimentally. Hence, it is always problematical whether experimental results gained under conditions at Time 1 will generalize to a different set of conditions at Time 2.

These methodological problems are well-known, at least to those familiar with social experiments. They have been addressed with impressive ingenuity and technical virtuosity in the various analyses and reanalyses of the negative income tax experiments, including those reported in this volume. On balance, though, the kind of precision the experiments have supplied may be of doubtful utility to policymakers. As the experiments have matured, a kind of infinite regress into methodological and theoretical complexity distances the results from the concerns of policymakers.

Imagine a member of Congress innocently asking, "How will an income guarantee, based on the principles of the negative income tax, affect the likelihood that its recipient will work, the amount of work that a recipient will do, or the amount of income that a recipient will earn?" An answer to this question that is consistent with the experimental evidence would have to be qualified in at least the following ways: (1) the nature of benefit packages available to experimental subjects was limited; (2) benefit packages available to control groups at the time of the experiment varied unsystematically from one setting to another; (3) the effects of an income guarantee on labor supply were different for different population groups within experimental settings; (4) different experiments produced quite different estimates of effects; (5) misreporting of income and work by experimental subjects influenced the results in ways that cannot be fully explained; (6) benefit packages available to poor people have changed since the experiments, making it difficult to estimate the effect and cost of changing the existing system; (7) little correspondence exists between the benefit packages tested in the experiment and the range of benefit packages that could be made available under a feasible reform proposal; and (8) the population of potential beneficiaries of an income guarantee would be significantly different from the population tested in the experiment.

To be sure, the answer could be simplified considerably by trimming off these methodological uncertainties, and by averaging results across experiments. To do so, however, sidesteps the methodological rationale for doing experiments in the first place. Being precise, it seems, also means being complex and equivocal. Methodological precision is apparently not the same as clarity or policy relevance.

Social science researchers are accustomed to these inward-turning spirals of ever-increasing complexity. This is the stuff of which social science reputations are made. The main product of research is, after all, proposals for more research. Elected officials may be forgiven, however, if they do not share this enthusiasm, since they operate under different time constraints and different incentives.

One of the major lessons of the negative income tax experiments, then, is that methodological precision is not positively related to the clarity or policy relevance of results. This is hardly a novel finding,[1] but it bears repeating before the promises for the next round of social experiments are made.

Political Issues

In addition to policy research, the negative income tax experiments were also political advocacy. The New Jersey Income Maintenance Experiment was conceived, in 1966-67, by a group of policy analysts in and around the Office of Economic Opportunity as a way of forcing the negative income tax into the political debate on income support.[2] To be sure, more politically astute ways exist of legitimizing a novel policy idea than to run a costly, long-term experiment with indeterminate results. But the advocates of the negative income tax were social scientists, not politicians, and they did what social scientists do best — they wrote a research proposal.

Over time, the experiments expanded and were accompanied by related research on poverty, spawning a significant analytic group within the Department of Health, Education, and Welfare and an enormous contract research industry outside the government. This analytic subgovernment became the main source of continuity in advocacy of the negative income tax during the welfare reform debates that followed.

Variants on the negative income tax found their way into the presidential or congressional arena on at least five occasions:

- In 1965 and 1966, the Office of Economic Opportunity presented President Johnson with proposals for a universal negative income tax. The proposals fell to budget pressures and the opposition of Undersecretary (later Secretary) Wilbur Cohen of the Department of Health, Education, and Welfare.[3]

- In 1969, President Nixon proposed the Family Assistance Plan, which combined an income guarantee with food stamps. The Plan passed the House in 1970, languished in the Senate, was reintroduced in 1971, failed again, and was abandoned by Nixon in 1972.[4]
- In 1974, Congresswoman Martha Griffiths, Democrat from Michigan, proposed an income maintenance plan consolidating several existing cash and in-kind programs. The proposal gained the endorsement of Griffiths' Subcommittee on Fiscal Policy of the Joint Economic Committee, and was the subject of hearings in both houses in 1976, but did not progress further.[5]
- At the same time as the Griffiths proposal, Health, Education, and Welfare Secretary Caspar Weinberger proposed an income-guarantee reform to President Gerald Ford, who rejected it in favor of more modest incremental changes.[6]
- In 1977, the Carter administration, after a period of bruising intramural combat between the Departments of Labor and Health, Education, and Welfare, introduced the Program for Better Jobs and Income which combined income supplements, earned income tax credits, and public jobs, in a supposedly zero-cost, comprehensive reform. The proposal never reached the floor of either house of Congress. The reasons offered for this failure included opposition from organized interests, conflicting Administration priorities, and budget constraints.[7]

The preparation of each of these proposals entailed considerable analytic staff work, which benefited, no doubt, from the existence of a well-staffed analytic subgovernment. In this sense, the federal government's investment in income support policy research paid off.

The published record shows, however, that the results of the negative income tax experiments entered the policy debate explicitly only twice. The first occasion was in early 1970, when the Office of Economic Opportunity, under pressure from Congress, released preliminary results from the New Jersey experiment less than two years after its commencement, showing what it claimed was a negligible impact of a negative income tax on labor supply. This report was immediately rebutted by a General Accounting Office study that labeled the results "premature."[8] The second occasion was in late 1978, when Senator Daniel Patrick Moynihan, Democrat from New York, announced in a speech on the Senate floor that evidence of high rates of family dissolution among recipients in the Seattle-Denver experiment had caused him to question his earlier advocacy of a negative income tax.[9] Neither of these occasions captures what policy researchers have in mind when they think about the use of research in policymaking. Both, however, demonstrate the political uses of evidence and the limited value of methodological sophistication in clarifying policy choices.

Another striking feature of political debate on income support is the relative infrequency with which empirical estimates of effects, of the type represented by the negative income tax experiments, figure in decisionmaking. Most of the debate on the Family Assistance Plan and on the Program for Better Jobs and Income, both within the executive and between the executive and Congress, focused on what might be called the architecture of proposals — the way various existing programs and their disparate benefit levels could be melded into a defensive reform package — and on the winners and losers created by various alternatives to the existing system. On occasion, the debate required estimates of population parameters and total costs. If these could be supplied by the analytic subgovernment, so much the better. If they could not, that was OK too — any reasonable estimate would suffice. In other words, policymakers were mainly concerned about the incremental effect of changes, on the architecture of benefit programs and on winners and losers. They were not overly constrained by the lack of solid causal evidence on the issues regarded as important to designers of the negative income tax — notably labor supply and family structure.

Judged in terms of their direct effect on policy, then, the influence of the experiments has been modest-to-negligible. But direct effects are deceptive. If we have learned nothing else from two decades of systematic research in the service of policymaking, it is that — to use Carol Weiss's terminology — the "decision-drive" model of policy research is less accurate descriptively than the "climate" model.[10] That is, research influences policy not by marshalling specific evidence in support of specific decisions, but rather by shaping policymakers' perceptions of the relevant problems and the feasible range of solutions.

This view tracks with John Kingdon's research on political decisionmaking, which concludes that "Academics, researchers, and consultants affect the alternatives more than the agenda, and affect long-term directions more than short-term outcomes."[11] In Kingdon's view, the formulation of a working agenda of policy problems, the formulation of alternative solutions to those problems, and the politics of decisionmaking normally operate as three separate, more or less autonomous processes. Only on rare occasions do these three streams converge into what Kingdon calls "policy windows," or opportunities for major changes in policy.[12]

Viewed from this perspective, the negative income tax experiments, and related policy research on poverty and income support, have had about the effect one would expect. The analytic subgovernment that grew up around the negative income tax is one of a few select locations for stockpiling policy options — or "option depots" — in the event that the problem and decisionmaking streams converge. This subgovernment has recently been joined by other, less social science oriented, more ideological option depots. When the problem-identifying and decisionmaking streams occasionally converge, as they have on four or five occasions over the past 20 years, these option depots supply some of the raw material for policy debate. This raw material gets unpacked and repackaged with other raw material in a variety of ways during the policy debate. The results are, for the most part, horrifying and depressing to the personnel of option depots, who see their careful work being defiled by amateurs and dilettantes.

This view also explains why the negative income tax experiments have had such a modest effect on income support policy and why the clear-cut vision of reform offered by their advocates is constantly messed up by policymakers. First is the issue of timing. The experiments ran on what might be called "social science time," or S-time. It has taken close to 20 years to get the experiments up and running, to accumulate and analyze the results, and to subject those results to the kind of critical scrutiny and secondary analysis required to tease out their strengths and limits. Even then, researchers have argued that the treatments were not in place long enough to give useful information about their long-term effects on labor supply. The policy process, on the other hand, runs on "political time," or P-time. The key determinants of P-time, for purposes of income support reform, are presidential elections and annual budget cycles. The calendar on P-time is shorter and more compressed than on S-time. On the few occasions when income support policy reform has surfaced in P-time, the results of the income support experiments have not been sufficiently mature in S-time to provide useful guidance to policymakers.

Second, the process of problem identification and political agenda-building contains much intelligence of importance to policymakers that is considered to be random noise by policy researchers. Advocates of the Family Assistance Plan and the Program for Better Jobs and Income, for example, consistently ignored signals from individual members of Congress and political advisers within the executive branch that "guaranteed income" and "negative income tax" were terms that carried very problematical overtones for certain key members and their constituencies. After the Family Assistance Plan proposal, the negative income tax became anathema to Al Ullman, Chairman of the House Ways and Means Committee, because of its extreme complexity, its potentially enormous cost implications, and its association with an earlier political debacle. Policy researchers found this aversion to be irrational and nonsensical, because after all, one could demonstrate analytically how existing benefit programs contained perverted and inferior versions of the same basic elements as the negative income tax. In other words, for policy analysts and researchers, welfare reform was the deliberate design of an income support system around certain policy variables with certain outcomes specified; for elected officials, welfare reform was the repackaging of programs so as to create a winning coalition. In the debate about the Program for Better Jobs and Income, analysts from the Department of Health, Education, and Welfare continued to serve up negative income tax-like options in the face of mounting evidence that key members couldn't stand them. The analysts were laboring in their own option depot, defending their product, disdainful of the difficulties their work was causing in the political arena.

Third, the process of political decisionmaking usually works in ways that are upside down, backwards, or perpendicular to the proposals of policy researchers. In the debates, for example, both President Nixon and President Carter stipulated certain budget and policy constraints early in the decisionmaking process that forced executive staff to make serious compromises in the architecture of the plans. The Family Assistance Plan contained a budget constraint and a stipulation that food stamps had to be treated separately and not cashed out. The Program for Better Jobs and Income contained a zero-cost constraint and the forced marriage of the Health, Education, and Welfare and Labor Departments through the amalgamation of cash assistance and jobs. In both these exercises, one could have produced a "better" plan — in both political and analytic terms — without the constraints imposed by presidential leadership, but the Presidents insisted on them.

Finally, policy research tends to focus, as it should, on the effects of options on recipients. In political decisionmaking, though, the relevant units of analysis also include organized interest groups and state and local governments. In both the Family Assistance Plan and the Program for Better Jobs and Income, the architecture of the benefit packages could be defended for recipients overall, but the plans collapsed under scrutiny of their effects on recipients on the margins of existing programs and their effects in high benefit states and localities. The intended beneficiaries of income support policy are not, for the most part, the interests that have the most influence in the formulation of that policy.

These conditions of political decisionmaking mean that large-scale policy research efforts of the type embodied in the negative income tax experiments will inevitably have a very limited influence on policy. The notion that social experiments could be used to leverage income support policy in the direction of a negative income tax has been proved by experience to be patently absurd. The problems of timing, or marshalling the right political intelligence for the right moment, and of building political constituencies around specific proposals far outrun the complexity of designing and running an experiment. The notion of deliberate investment in option depots and of the careful stockpiling of evidence on the complex and equivocal effects of reform policies is not absurd, however, if it is accompanied by low expectations that these options and evidence will have a direct effect on policy.

---

[1]Cohen, David and Michael Garet, "Reforming Educational Policy with Applied Research," *Harvard Educational Review*, 45 (1975) pp. 17-43; Lindblom, Charles and David Cohen, *Usable Knowledge*, New Haven, CT: Yale University Press, 1979.

[2]Williams, Walter, *The Struggle for a Negative Income Tax*, Seattle, Washington: University of Washington, Institute of Government Research, 1972, pp. 2-11.

[3]Ibid., pp. 4-5.

[4]Moynihan, Daniel Patrick, *The Politics of a Guaranteed Income*, New York: Random House, 1973; Lynn, Laurence and David Whitman, *The President as Policymaker: Jimmy Carter and Welfare Reform*, Philadelphia, PA: Temple University Press, 1981, pp. 18-34.

[5]Lynn and Whitman, pp. 42-43.

[6]Ibid.

[7]Ibid., passim.

[8]Williams, op. cit., pp. 23ff.; Moynihan, op. cit., pp. 191ff.

[9]Lynn and Whitman, op. cit., pp. 247ff.

[10]Weiss, Carol, "Research for Policy's Sake: The Enlightenment Function of Social Research," *Policy Analysis*, 3 (1979), pp. 531-545; "Improving the Linkage Between Social Research and Public Policy," in Laurence Lynn, ed., *Knowledge and Policy: The Uncertain Connection*, Washington, DC: National Academy of Sciences, 1978, pp. 23-81.

[11]Kingdon, John, *Agendas, Alternatives, and Public Policies*, Boston: Little, Brown, 1984, p. 71.

[12]Ibid., pp. 92-93 and 173ff.

Discussion

*Robert D. Reischauer**

Welfare reform has always presented policymakers with two related, but distinct, problems. The first has been the technical problem of designing a reform package that could achieve the desired results within the budget constraint. The second has been the political problem of creating a constituency that could get the reform package enacted.

This second problem exists because, unlike farm, housing or education policy, welfare policy has been neither an area popular with the public nor an area that provided much political payoff for politicians. The potential beneficiaries of welfare reform have never been politically active and, therefore, have had little clout in Washington; and the public has held a generally hostile view of welfare programs. When politicians have supported welfare reforms, they have not done so because they anticipated PAC contributions from welfare rights organizations or long lines of appreciative recipients at the voting booths. Rather they have supported such policies because they thought they were the right thing to do and because they concluded that the reforms would not generate a backlash from the voters.

Given the hostile political environment that exists for welfare reform, an interesting policy question is whether the income maintenance experiments could have helped to build a constituency for reform. Could the experiments have reduced the public's distaste for radical reform or mobilized the low-income population into an effective interest group in behalf of change?

The simple answer to these questions is "no." In fact, there were good reasons to expect that the experiments would, if anything, strengthen the hand of those opposed to reform. Part of the reason for

*Senior Fellow, The Brookings Institution.

this negative conclusion lies in the nature of the negative income tax and part lies in the particular focus of the income maintenance experiments.

Richard Elmore has argued that the political objective of the experiments was to legitimate the negative income tax concept — to give it standing with policymakers. An essential first step in this process was for the Office of Economic Opportunity technocrats to convince the White House and congressional policymakers that a negative income tax was a technically feasible policy option. The income maintenance experiments played a modest role in doing this. While the results of the experiments may have entered the public record of the policy debate only twice, their contribution was more significant than Elmore suggests. The findings of the experiments were discussed at dozens of planning and strategy meetings between congressional advocates of welfare reform and policy formulators in the executive branch. They influenced the design of the Carter welfare reform plan in numerous ways.

However, the major political hurdle was not to convince the policymakers that a negative income tax was technically feasible but rather to convince the American public that this radical approach to the poverty problem was acceptable. And the experiments were not capable of doing this, primarily because the negative income tax was designed to address the deficiencies that the policy elite saw in the existing welfare system, not the shortcomings that most concerned the general public. The public felt that welfare costs were too high and that caseloads were expanding too rapidly; they imagined that undeserving freeloaders who were capable of work were weaseling their way onto the rolls in increasing numbers; they feared that the system was encouraging marital instability; they suspected that cash assistance was being squandered on booze, color TVs and other unnecessary expenditures; and they felt that welfare was creating a permanent dependent class. The policy elite, in contrast, was more concerned with the inequities created by interstate differences in payment levels and eligibility requirements, the failure to provide assistance to two-parent families, the general inadequacy of benefits, the stigma associated with receipt of welfare, and the political contentiousness that surrounded discussions of income redistribution policy.

On many dimensions, a negative income tax was bound to exacerbate the very aspects of the existing welfare system that most concerned the public. For example, a negative income tax would increase costs and welfare rolls, provide assistance to more able-bodied adults, reduce popular in-kind benefits in favor of more cash assistance, and allow families to remain on welfare for life. In addition, it should be noted that a true negative income tax would cut away what little there was in the way of a mainstream political constituency for welfare programs, the providers of in-kind benefits: the farmers who benefited from the food stamp program; the real estate interests who gained from the various housing subsidy programs; the professionals who provided the various social services; and so on. Overall, a negative income tax clearly went against the popular grain and, therefore, would be difficult to legitimate.

Given this situation, the political task for the experiments was to provide information to convince the public that they would like the results of the new system even if they did not like the manner in which a negative income tax would provide assistance. In other words, the experiments would have to highlight those aspects or responses to a negative income tax that might attract the support of the nonrecipient public. This strategy would have suggested emphasizing its impact on family stability, crime and delinquency, health status, nutrition, school achievement, and other effects that would indicate that the recipients of a negative income tax ultimately could attain self-sufficiency.

While the income maintenance experiments did examine such responses, these dimensions were of secondary interest. The experiments focused on the measurement of the labor supply responses to a negative income tax. Inevitably, this focus made the political problem worse because any negative labor response would constitute a political liability. While the policy experts might have been overjoyed to find that the reduction in labor supply was small, a skeptical public and congressional critics would not differentiate between a 5 percent and a 20 percent reduction in labor supply. In both cases, more indolence was being rewarded at the taxpayer's expense.

The experiments were not only incapable of allaying the public's apprehensions concerning radical welfare reform, they were also not capable of building a low-income constituency for reform. Because the experiments involved only a tiny fraction of the low-income population, most of the potential recipients of a negative income tax knew nothing about the options that were being tested. Moreover, to the extent that politicians used the experimental results to generate support for a negative income tax among low-income groups, they ran the risk of alienating the general public. A reformed system that reduced stigma, intruded less into the lives of the poor, provided more generous benefits, and offered assistance to families that previously had been ineligible might rally the low-income population, but would lose the middle class.

The general conclusion that arises from examining the experience of the income maintenance experiments is that, under most circumstances, social experiments have very limited political utility. Policy analysts may find them a useful way to convince politicians that a certain policy is technically feasible. But in doing so they are likely to focus attention on the politically unpopular behavioral responses to the proposed policy.

Imagine what might have happened if a social experiment had been mounted in the 1920s to test the feasibility of a government-subsidized old-age pension system. The results from this experiment would have revealed that a large proportion of the elderly would drop out of the labor force under the proposed pension program. While some of these retirees would have health problems that made it difficult for them to meet the physical demands of their jobs, many would be quite capable of continued labor force participation. The experiment's results would also have shown that the proposed program would induce some workers to drop out of the labor force even before they were eligible to receive a pension. Would this be rewarding indolence?

Careful analysis of the experiment's effects would also have revealed that the new program would lead to a reduction in private retirement saving and a slight tendency for employers to cut back on their pension programs. In other words, only a portion of the federal pension payment would represent a net increase in the living standards of the retirees.

The sociologists and psychologists analyzing the results of the experiment would have discovered that the proposed program would threaten accepted family patterns. Many of the affected elderly would choose to move out of their children's homes, some to small squalid apartments. A few of the elderly would even move to distant places like Florida where their children could not care for them. By all measures, the proposed program would lead to a reduction in contacts between parents and their children and a decrease in the sense of responsibility that children would feel for their aged parents.

These results, while not surprising, would have represented a political liability for advocates of social security because they would have clashed with prevailing values and behavioral norms. But major social reforms, like social security or a negative income tax, inevitably will change prevailing values, behavioral norms and the political environment. In prospect, many of these changes may appear threatening and therefore undesirable. After some years, however, they will become not only accepted but also desirable attributes of modern living. For this reason, advocates of major social policy changes should think twice about the desirability of experimentation. What is gained in the way of an understanding of the micro effects of a proposed policy may be lost in the political realm.

An Economist's View of the Income Maintenance Experiments

*Robert M. Solow**

I am cast on this panel as offering an economist's view of the policy lessons to be drawn from the income maintenance experiments. That will be true enough if you take the word "an" seriously. I am indubitably a card-carrying economist. But I have the feeling that — on these particular matters at least — my views are not always those of the typical economist. Part of the difference, but only part, arises because I am primarily a macroeconomist, so I look at the questions now at hand as a partial outsider.

Any time you are about to utter heresy it is a good idea to quote some highly respectable authority. So let me remind you of something John Stuart Mill wrote in the Preface to the *Principles of Political Economy.* "Except on matters of mere detail, there are perhaps no practical questions, even among those which approach nearest to the character of purely economical questions, which admit of being decided on economical premises alone." Actually, I intend to go even further than Mill: there are very few analytical questions about mass human behavior which admit of being decided on economical premises alone. And those few are not the ones we are dealing with at this conference.

The formal purpose of social experiments is to provide knowledge about mass human behavior that will be useful in the design of policies. So the lessons for policy depend on what we read the experiments as saying about behavior. In commenting on that issue, I am drawing also on my experience with other social experiments, in particular those dealing with supported work and with various work-welfare schemes, designed, conducted and analyzed by the Manpower Demonstration Research Corporation.

---

* Professor of Economics, Massachusetts Institute of Technology.

The surest generalization that emerges from all such experiments is that the implications for mass human behavior are weak. That does not mean they are not important; but they are uniformly weak. I use the non-technical word "weak" to cover a couple of distinct characteristics. In the first place, statistically significant response-coefficients are hard to come by. If there is any signal in the experimental results, it is rarely audible above the noise. This is hardly surprising; it is a very common outcome in cross-section studies with individuals as the unit of observation. No doubt it reflects both the inherent variability of each individual's behavior and the variation among individuals in their average response above and beyond what can possibly be related to observed and observable characteristics. Research workers usually get their kicks from large t- and F-statistics, so this general lack of statistical significance is usually a disappointment; but I think it has its policy uses, as you will see. It might sometimes be possible to design an experiment for adequate precision or power at some favored point in parameter space. But that is hardly ever what we really want.

Even when a statistically significant response-coefficient surfaces it is usually small. The elasticity of this with respect to that is rarely large for the thises and thats we are concerned with. To take two examples: (1) Gary Burtless shows convincingly that the labor-supply response to a realistic negative income tax is in the expected direction but fairly small; (2) in the supported-work experiments, even when a favorable post-program response-coefficient could be estimated, as with the sub-sample of mothers receiving aid to families with dependent children, it was far from dramatic. Gains in employment and earnings were small, but definite. I could document an analogous statement about the work-welfare experiments Manpower Demonstration Research Corporation is now doing. We are so accustomed to this sort of outcome that our first reaction to a large statistically significant elasticity is to say: Wait a minute; this must be spurious, the result of a misspecification. And almost always we can convince ourselves that this skeptical reaction is correct. I don't find this sort of outcome terribly discouraging either. We are talking about fairly commonplace aspects of behavior, not about responses to exotic stimuli or extreme situations. If sharp responses were to be expected we would already know about them; nobody spends millions of dollars to verify the obvious. (It goes without saying that sometimes common knowledge will be wrong and the obvious is not only not obvious, but false. That's life, but it does not undermine the reason I gave for expecting small responses.)

If I am right about this, that the typical outcome of a social experiment is some weak conclusions about response elasticities, it has an important implication. The prevalence of small effects opens the way to alternative interpretations of the research findings. The interpretation adopted will depend a lot on the interpreter's ideological and doctrinal preconceptions and only a little on the detailed experimental results themselves. It is the same principle that governs those personality tests in which you are shown a picture of an ambiguous scene and different people, interpreting the same picture, will tell you the story of their own lives. Thus I think Gary Burtless hit the nail right on the head when he concluded his survey of labor-supply responses by observing that policymakers "seem far more impressed by our certainty that the efficiency price of redistribution is positive than they are by the equally persuasive evidence that the price is small." Because the price is small and not well-defined, policymakers can find what they are looking for. We all know what they are looking for these days.

One obvious reason for the prevalence of weak results is Mill's dictum. These experiments do not take place in a test-tube and they do not involve identical individuals. There is just a lot more going on than can possibly be controlled. And many of those things are not even economical at all. I thought of this especially while reading Glen Cain's meticulous and carefully inconclusive evaluation of the well-known findings about marital instability. Of course current and prospective income under various circumstances is one of the forces pushing a family one way or another. But there are many others, probably more acute, impossible to measure and control for, and — this is the important point — very likely correlated with some of the things we do measure, but in complicated ways. Since those unmeasured forces necessarily get parked in the noise, it is no wonder that clear, comprehensible, and robust response-coefficients are hard to find in social experiments.

If the experiments usually offer little basis for discrimination among quite different interpretations of the outcome, do they therefore offer no lessons for policy? No, I don't think that is the right conclusion to draw. I think that the income maintenance experiments and others like them tell us a lot about social policy, though maybe not the things we are accustomed to look for. Once again it is Mill's dictum that points the way.

Why was there, back in the 1960s and 1970s, a brief flicker of interest in the negative income tax, intense enough to give rise to expensive experiments? It is hard to think back across the ethical desert that is American national government today, but I think I remember some of the reasons. There was a feeling that we were at last in a position to eliminate poverty, that it was the right thing to do, and that the direct way to do it was to transfer income directly to people who would otherwise be very poor. This was combined with a feeling that the existing hodgepodge of categorical transfer programs involved the bureaucracy deeply, meanly, and inefficiently in running the lives of the recipients. In particular, it was thought that the rules governing AFDC had the effect of inducing two-parent families to split up. This strand could be thought to fall under the second category of reasons, but it was prominent enough to be worth independent status. The negative income tax seemed to be a way of fixing all of these things at once, and doing it through a bureaucracy with which *everybody* was involved. It attracted people at both ends of the normal political spectrum. This may have been very important at the time. After all, we didn't get a negative income tax, just some experiments. The unusual sight of Milton Friedman and James Tobin agreeing (but not really) about something could help bring that about. I think Wildavsky's categorization of the political spectrum is too *ad hoc* to be fundamentally useful; but I also think he captures a lot of what was going on in this particular episode.

The one possible hitch was the fear that a decent guaranteed income level, combined with the high tax rates necessary to keep from transferring a lot of money to people above the poverty line, would induce many recipients to withdraw from work. That outcome would fit badly with a very important strand of the American ethic. The experiments were designed primarily to test that possibility. (At least so it seems to me. I may here be exhibiting the economist's occupational bias.)

If that is what the experiments were about, they did provide something of an answer. There is a labor supply effect, as every economist thought there would be; but it could hardly be described as large enough to jeopardize the work ethic. Besides, the inducement to withdraw from the labor force seemed to be stronger among women than among men; it might easily have weakened over time as women generally have become increasingly involved in the world of work. A culture that can pat itself on the back unceasingly after having gone seven years with an unemployment rate higher than 7 percent can hardly complain that its foundations are being eroded by so small a withdrawal from the labor force. There is no particular reason to suppose that a return of those lost souls to the labor force would have increased *employment* perceptibly under current conditions.

The possible effects on marital instability are some cause for alarm, though Glen Cain makes one wonder if they are real. No one would want the transfer mechanism to contribute to the breakup of marriages. By the way, is there any reason to connect any measured increase in marital splits with the negative income tax *mechanism* itself, as distinct from the change in current and expected disposable income? If the couples that did split had encountered the same income possibilities in the private market, would they have split up? There is no merit in insisting that the transfer mechanism keep couples together who want to break up for other reasons, and do so by inflicting unnecessary poverty on them. The negative income tax experiments may not have been causing anything, just telling us what was already there, but suppressed.

Maybe this view of the experiments provides a model for social experiments in general. Suppose society wants to do something because it is the right thing to do, not for purely economical reasons. There will sometimes be economical doubts, worries that doing the right thing might be unsuspectedly costly. A well-designed experiment can help find out, and the prevalence of weak results is not an obstacle. It only means that a lot of the time the experiment will tell us that, for all the reasons discussed earlier, the kinds of changes we contemplate will not turn the world upside down. Many of the most important things people do they do for reasons that have little to do with the price mechanism. I think that the great shift in educated opinion, away from belief in income maintenance-type solutions toward belief in multi-part tailored, work-related programs, had little to do with the outcome of the negative income tax experiments. The generic fact that results are weak merely permits the social consensus to work itself out and to convert the research community without undue strain on its conscience. It must also be remembered that the nature of the poverty problem seems to have changed during 20 years and that should change the preferred solutions.

I should add that experiments have the added advantage of providing some information about the system's capacity to administer and operate a new policy mechanism. This is important; it is something that the Manpower Demonstration Research Corporation has always paid a lot of attention to in its experiments and demonstrations.

The main disadvantage of the social experiment as a policy tool is that it may often leave us having to explain to ourselves why we do not do the right thing, when it is costly but not terribly costly. But that is not a serious problem either. One of the things Americans are best at is kidding themselves along.

Discussion

*Edward M. Gramlich**

I would have thought that this conference would commemorate the income maintenance experiments, but in fact the experiments have taken quite a beating. Yesterday Arnold Zellner made extensive criticisms of their statistical properties. Today Lee Rainwater criticized their initial conception, method, and policy interpretation. Elmore found the notion that policy experiments of this sort could be used to leverage policy support patently absurd. Nathan feels that neither the negative income tax nor the experiments with it were well-advised. Robert Solow's paper was also critical of the experiments. When he wrote it, he probably thought that his paper was being too critical, and he was quite apologetic. But compared to the others, Solow's paper is mild. To quote another famous social philosopher (Solow himself, in his 1974 Ely lecture), he must have felt ''like a nice independent rat, trotting down to the sea, and suddenly discovering that he's a lemming.''

Solow is only rarely a lemming, a feeling I have much more experience with. Today, however, I want to try to play the rat and stick up for the experiments. I begin with a standard Robert Hall proposed yesterday when he suggested that the experiments be compared to what the government was doing at the time. Hall compared the experiments to the Post Office, which seems like a cheap shot. But we could compare the experiments to other major social experiment efforts of the Office of Economic Opportunity (OEO) at the time. There were three altogether. One, on educational performace contracting, tested a policy that soon became viewed as a failure, using an experiment that was very flawed but ultimately arrived at useful ''nail in the coffin'' type data because of

*Acting Director, Congressional Budget Office, and Professor of Economics and Public Policy, University of Michigan.

the well-known inadvertent finding. That finding was that it became very hard for OEO to litigate the performance contracting claims with private companies, raising the ugly spectre of small school districts trying to fight it out with the legal departments of national teaching companies. Another, for educational vouchers, never even got off the ground because OEO could not find any school boards interested in conducting an experiment. The third, the negative income tax, was at the time viewed as the jewel in OEO's crown. The experiment was working, and the policy was interesting to many people, even some who were not economists. In this sense, while we can look back, rub our beards, and pronounce the negative income tax a failure, we should give the experiment its due — unlike the others, it at least lived on to get roasted 20 years later.

But the experiment did more than that. There are basically two reasons why many are now pronouncing the experiment a failure:

1) the research was inconclusive;
2) interest in the policy under investigation waned.

Yesterday we discussed the first reason extensively, and I take it by now the consensus is that not *all* the research was inconclusive, and that it is pretty hard even to imagine what conclusive research would be on matters such as marital splits. Today we are talking about the second issue, one on which I would like to spend most of my time.

The reason why interest in the negative income tax as a policy option has waned also brings up Solow's remark about the "ethical desert that is the American national government today." (Since I am now part of that national government, this remark does put me in an awkward position.) Back in 1970, and probably also today, Solow would have voted for a pure negative income tax without any work requirement because "it was the right thing to do." A decent negative income tax would cost about $20 billion, $100 per capita, about one-fifth of what the average American spends on voluntary charity these days. Most people give to charity without any effective monitoring for leaky buckets, and one would think we could do the same for income transfers. Solow appears to adopt this line of reasoning when he calls income maintenance the right thing to do. I certainly agreed with him back in 1970, and would still today but for another problem that I was not then very sensitive to: unfortunately, a pretty basic problem.

Theorists rationalize the existence of income transfers in the first place on the basis of taxpayer-donors' altruistic motive. But any altruist knows that one important thing to worry about is free riding: if there is some other altruist willing to make support payments, an altruist can free ride, see the donation given, and save his money.

It takes just a slight extension of this logic to realize that with income support payments, there is a potential free rider — the recipient himself.

It may be that free riding is minor, that it is not costly to the donor, and that leaks in the bucket are modest. But if free riding exists at all, it can be tremendously destructive. The donor-taxpayer is working a little bit harder to support a poor person, and here we have the poor person working a little bit less. Any donor-taxpayer, even a very ethical one, is likely to become so enraged by such an outcome that he may cut off all support. And even if he does not, his democratically elected representative might be inclined to. And even if that representative does not, he may have to run against an opponent who could make hay out of the issue. When all is said and done, income support under those terms will probably be very limited.

This, I think, is what blew the pure negative income tax out of the water politically. Unless donors could *assure* themselves that responsibility for supporting recipients would be shared by recipients themselves, in the form of work requirements, child support enforcement, and other things that sounded punitive back in 1970, they are simply not interested. And it is not even unethical for them to be uninterested. This is why, I think, Nathan tells us that politicians did not like the pure negative income tax even in 1970, and Burtless tells us that politicians are more impressed by the positive price than by the small price. And this is why I also think we should stop talking about ethical deserts. They may not be ethical deserts after all, and whatever they are, they have been around for a while.

I should point out that by responsibility-sharing, I do mean sharing. Unlike Charles Murray, I do not think that poor people need to provide *all* of their own support. There can be altruism, and it can be highly satisfying to donors as long as donors perceive that poor people are doing their share. And such a feeling is probably best for the long-run self-esteem of the poor too, another important value that should not be ignored.

Unlike all pessimists who are down on both the negative income tax and experiments with it, this notion of responsibility-sharing shows me a silver lining in all these gray clouds surrounding the world of income support these days. It is just possible that national policy, nudged along by state governments (the true social laboratories) and even unwittingly by the Reagan administration (mainly in the form of enabling legislation passed in 1981), is evolving a successful income support strategy that contains heavy doses of responsibility-sharing. States are more and more requiring welfare recipients to work or search for jobs as a condition of getting benefits. They are also enforcing child support obligations on absent fathers. Both make sense in terms of responsibility-sharing; the first may even pass the Manpower Demonstration Research Corporation's benefit-cost test, and evidence indicates that it is viewed as fair by recipients.

Once the perception that welfare involves responsibility-sharing takes hold, we may even see governors begin to bring welfare out of the closet and brag about their "humane and responsible" approaches to the problem. One such governor is in the state just south of us, Dukakis of Massachusetts. And with governors and other politicians talking up this new view of public assistance, these programs may begin to grow in popularity, and real benefit levels may even stop losing ground.

There is a silver lining for social experiments too. The Manpower Demonstration Research Corporation's interesting and successful evaluations of employment for welfare recipients have somehow managed to take root in this desert. They have benefited technically and administratively from earlier experiments with the negative income tax, and they now have the huge political advantage that their policies under investigation seem to be popular and have stayed so, seemingly because they have worked out the sharing of responsibility more satisfactorily. To respond to those who proposed retrospectively a resident anthropologist, we now even have at least one new large-scale project, by Bill Wilson at Chicago, that combines survey and ethnographic work. Hence for all the mistakes made by the earlier generation of social experimenters, and all the bad political luck that was suffered, we might even be getting another chance. Let's hope things work out better this second time around.

Views of a Policymaker and Public Administrator

*Barbara B. Blum**

This paper, which considers the income maintenance experiments from a welfare administrator's point of view, explores two major questions. The first is one of process: What is the relationship between the way in which the experiments were conducted and their reception by welfare officials? The second question concerns substance: What lessons for administering today's welfare system are suggested by the goals with which the experiments were undertaken and by the knowledge they generated?

Before addressing these questions, however, it may be helpful to briefly set them in the context of the major papers before this conference. A review of the papers by Burtless, Cain and Hanushek highlights the diversity of environments studied in the negative income tax experiments — variations in the demographics of the sites, in their grant levels, their economic conditions and their tax structures — a diversity that mirrors the heterogeneous circumstances of the poor in this country. Such diversity heightens the difficulty of any effort to reform the nation's income maintenance system. Moreover, the variation of environments studied was compounded by the wide variation of experimental designs, which encompassed so many different benefit and tax levels, research samples and services. Thus, in reflecting on how these experiments might have affected welfare administrators in the past and how their lessons might illuminate lessons for the future, it is useful at the outset to recognize that these experiments were ambitious, highly complicated and probably overly elaborate in design. While this paper does not focus directly on these characteristics, it will have occa-

*President, Foundation for Child Development, and former President, Manpower Demonstration Research Corporation.

sion to touch on the question of their implications for welfare administrators and welfare reform.

The Conduct of the Evaluations

At the beginning of my own tenure as New York State Commissioner of Social Services in 1977, research on the income maintenance experiments was drawing to a close. Over my five years in that position I was only very generally aware of the negative income tax experiments. Extended discussion of the study, either within my own department or among my colleagues in other states, did not occur. My relationship to the experiments seems to have been typical of that of most welfare officials, both at this stage of the evaluation and earlier. A scan of the 1968-78 issues of *Public Welfare*, the major journal for administrators published by the American Public Welfare Association, suggests that this premise is correct.

In a 1974 article on "The Current Status of Human Services," Mitchell Ginsberg and Norman Lourie referred to the New Jersey experiment along with other research on the welfare system as evidence that transfer payments have "little or no effect on the willingness to work and work incentives" but also cautioned that "it would be naive to assume that basic public policy decisions will be made primarily or even substantially on the basis of research findings."[1] Other issues of the journal published over this 10-year period contained a half dozen additional brief discussions of the concept of the negative income tax, mostly raised in connection with the proposed legislation on the Family Assistance Plan, but no comment on the negative income tax experiments themselves is apparent.

Did the negative income tax develop a higher profile among welfare administrators in the communities where research was conducted? Because this question was not of primary interest to almost anyone who followed the experiments, it is difficult to piece together the answer. There is, however, some evidence that speaks to this relationship.

In New Jersey, there was no official connection between the state and the experiment. Contact between welfare officials and researchers at that site was apparently limited to a controversy over whether researchers were obliged to release their records to the welfare department so that officials there could determine if some individuals were collecting payments under both programs.[2]

To prevent a recurrence of such a conflict, the relevant states were officially a party to the experiments in both Gary and Seattle-Denver. As Director of the Office of Income Maintenance Experiments in the State of Washington's welfare department, Joseph Bell reports that he was able to work closely with welfare officials on tracking clients to be certain that they were not receiving payments from both aid to families with dependent children (AFDC) and the experiment. Also, some officials indicated that something could be learned from the administrative innovations of the experiments. However, administrators seemed to be only mildly interested in the possibility that a negative income tax might replace the welfare system as they knew it; the attitude seems to have been that by and large academic research would have little effect on programs in the real world.

In Gary, Indiana the experiment — in the words of its research director, Kenneth Kehrer — "in effect set up its own welfare system," taking over the administration of the cases of participants who entered the negative income tax experiment as recipients of AFDC benefits or food stamps.[3] Because the caseworkers for these individuals were employees of the county welfare department working under subcontract, there was by definition a close connection between the experiment and the department. As in Seattle, however, involvement of welfare workers and officials was largely limited to the "nuts and bolts" administrative issues.

For at least two reasons, it is not surprising that except for the aspects of the experiments that directly concerned them, welfare officials were generally out of touch with the progress and development of the evaluation. First, like almost all long-term research, the negative income tax experiments operated within a different time frame than the one to which administrators and elected officials must usually respond. The experiments were expected to take a number of years to produce answers about how to improve the nation's income maintenance system. Meanwhile, the system itself was changing. In the turbulent 10-year period between the late 1960s and 1970s, welfare administrators confronted a variety of new and pressing issues — the separation of services from income maintenance functions within the AFDC program; the need to relate to many new programs including Medicaid, supplemental security income (SSI), Title XX and a greatly expanded food stamp program; and demands that more attention be devoted to error rates and to work obligations for welfare recipients. Given the pressures to adapt to these changed circumstances, it is understandable that administrators did not focus on the fine points of an experiment that they perceived as offering few solutions to immediate problems.

In connection with the subject of different time frames, it should be borne in mind that by definition an extended research project outlives the terms of office of most political leaders and officials. In New York State, for example, the period of the negative income tax research covered the terms of three governors and five welfare commissioners. This natural rate of turnover means that officials within a political system of bureaucracy have to be actively engaged if they are to devote attention to the unfolding of long-term research.

The difference between the time perspectives of administrators and researchers is apt to create obstacles to sustained communication between the two groups in almost any long-term evaluation. A second reason for the existence of such a gulf in the case of the negative income tax research is more specific to these particular experiments and the questions they asked.

The experiments were designed to determine whether it was possible to replace existing programs with a radically new and improved income maintenance system. With this approach to reform, the researchers had little incentive to establish regular channels of communication — except for those they specifically needed to do their work — with welfare administrators. Had the experimental results been strikingly positive and subsequently translated into policy, these administrators would have been called upon to function in a very changed system — or, more likely, they would have been displaced. To the extent that the researchers expected such an outcome, they may well have concluded that there was little to gain from interchange with officials of the existing programs.

The vision that led the designers of the negative income tax experiments to develop a new program model was that of an income maintenance system that was more uniform, rational, and fair and less intrusive. The experiments provide only limited information about what might have been achieved with a permanent negative income tax. We do know, however, that despite the numerous changes that have taken place since the inception of the experiments, the current system still falls far short of the ideals that inspired them. The problems in the current AFDC program suggest that policymakers interested in replacing it with a negative income tax had ample justification for examining the possibilities of wholesale reform. Nevertheless, there are trade-offs involved in trying to build a new system from the ground up — or even just in testing an innovation *outside* of the system where it must ultimately operate — as opposed to pursuing change within an existing institution.

Two examples from the work of the Manpower Demonstration Research Corporation (MDRC) may help to illustrate the point. MDRC was formed in 1974 by a consortium of federal agencies to evaluate the National Supported Work Demonstration, a structured work experience program for hard-to-employ individuals. At the end of a five-year research period, the program was found to be most effective in improving the employment prospects of long-term welfare recipients.[4] However, with its local projects either developed specifically for the demonstration or operated by social service agencies in the community, the project forged no lasting connections to the AFDC system.

Over the years since the experiments were completed, policymakers have expressed considerable interest in the Supported Work results, and the strategy has been subsequently tried in scattered locations around the country. However, there has been no widespread adaptation of the technique for use with the welfare population. Undoubtedly this is largely because in an era of fiscal restraint, this program, even though ultimately cost-effective, requires a relatively large upfront social investment. Nevertheless, it may be that the demonstration's institutional distance from the welfare system further decreased the likelihood that Supported Work would be widely used within that system.

In 1982 MDRC undertook another evaluation centering on employment for welfare recipients, the National Demonstration of State Work/Welfare Initiatives, which is now just past its midpoint. The Demonstration is examining programs for welfare recipients in 11 states that, in response to the new flexibility offered them under the Omnibus Budget Reconciliation Act of 1981, have chosen to develop their own strategies for alleviating welfare dependency — centering on job search and unpaid work experience with some education and training. In contrast to the Supported Work demonstration, the Work/Welfare project consists of a series of self-contained tests, each with its own schedule, but from which the researchers draw general conclusions. Thus, in a practice that might have been useful for the income maintenance studies, cross-site analysis has been built into this evaluation from the outset.

The Work/Welfare Demonstration has been intentionally operated with ties to the welfare system. For example, in addition to keeping state and local welfare officials from all states in the demonstration abreast of the findings on their own programs as they emerge, MDRC has periodically brought them together to exchange views and insights about their programs and to discuss the broader implications of the findings. In another undertaking that might have benefited the negative income tax experiments, the interim demonstration findings have been communicated to these officials and to others, particularly elected officials, in a brief and relatively nontechnical summary document.[5] Welfare officials in participating states have cooperated in protecting the integrity of random assignment and have otherwise facilitated the progress of the research. Administrators, governors, legislators and congressional representatives have evinced a sustained, and often keen, interest in the findings, and generally seem to have grown in their understanding of the contribution that experimental research can make to their programs.

One very important facet of the demonstration is that the individual programs under study have been developed by the states themselves, which therefore have a vested interest in learning about the strengths and weaknesses of the new model. This decentralized approach to the formulation of research questions differs from the practice in many experiments, where the program model is first developed by an outside authority, and then state or local departments or agencies are invited to take part in a test of its effectiveness. For example, Kenneth Kehrer recalls that in the early 1970s the State of Indiana approached the Department of Health, Education, and Welfare with a proposal to test a social services voucher plan, and that the Department instead prevailed upon the state to accept the income maintenance study that ultimately became the Gary negative income tax experiment.

Given the ties between the Work/Welfare Demonstration and its participating states, there is likely to be a more direct and clear line of influence from its major findings to welfare programs than was the case for Supported Work, and certainly for the income maintenance experiments. Even at this point in the evaluation, states are already shaping and refining their programs in terms of what they have learned from the research.

This does not mean, of course, that the Work/Welfare model of fielding an experiment is the only appropriate one. One of the reasons that the Work/Welfare states and their welfare agencies have been as susceptible to influence as they have is that the innovations tested in the Demonstration — the imposition on recipients of largely mandatory work-related obligations such as job search and workfare — are changes that can be accommodated without a major overhaul of the current system. They are also policies that states can hope to pursue even with the limited resources currently available to them for social programming.

There may also be — as seems to have been the case at the outset of the negative income tax experiments — good reason to study more sweeping proposals for reform. However, those who undertake such research should recognize that one disadvantage of a less incremental approach is that it is much more difficult to work closely with officials in the current system to jointly identify and introduce the changes suggested by research results.

Notwithstanding the many forces that kept administrators and other officials from becoming more closely involved in framing the negative income tax research questions or exploring the results as they emerged, it is instructive to speculate briefly on what might have transpired if this had occurred. One cannot know, but it seems probable that as the research took shape and progressed, these officials might have wanted to probe more closely into some of the intriguing questions that remain even today when we review the results. They might have, for example, focused on the concept that Burtless refers to — in the somewhat otherworldly terminology of theoretical economics — as a "sale on leisure."[6]

They might have asked if the ''leisure'' could be an opportunity for a single parent to make a modest reduction in work or earnings that could, in turn, have a positive impact on his or her children's school attendance and performance. They might have asked more about the interrelationships between the ''nonlabor'' aspects of the experiments discussed in Erik Hanushek's paper.[7] They might have tried to push hard against the findings on marital stability and family composition to find out whether these results really could offer guidance on the design of income maintenance programs, or if they were too limited to do so.

Would such attention and questioning from people involved in the real-world programs and policies have produced better answers? Perhaps no. Non-researchers can be impatient and unforgiving about the difficulties of using data to speak to complicated policy issues, and the three major papers presented to this conference attest to the many difficulties of this nature that arose in the income maintenance experiments. Still, one cannot help but wonder if over the decade in which the negative income tax was studied, a steady infusion of such interest would have given these questions a little more urgency, moving them closer to the center of public awareness. Citizens and politicians might have concluded that if these concerns could not be addressed by a negative income tax, they at least deserved more attention. In brief, we cannot be certain what cross-fertilization of research with the interests of public officials would have yielded. Perhaps the results would have been only interference, but perhaps they would have brought forth a process whereby issues posed in the experiments found their way more rapidly into public discourse.

Administrative Lessons from the Experiments

While it is true that the major findings from the negative income tax experiments had little or no direct impact on the welfare system, there was an unanticipated spillover from another aspect of these studies. In light of the fact that the association between researchers and welfare officials centered on administrative matters, it is perhaps not surprising it was in the area of program procedures that the negative income tax experiments left their strongest mark on the system.

To gather information on the income and employment status of participants, the experiments developed practices that differed in three important ways from those in effect at most welfare offices during this period. First, rather than an infrequent face-to-face redetermination, participants were generally required to report on these matters on a monthly basis by completing and mailing in a form. Second, these data were processed automatically. Third, in an innovation known as retrospective budgeting, the biweekly benefits sent to a family were based on their circumstances for the previous month, not on what it was anticipated they would need for the next one.

The reaction to these innovations was generally positive. In New Jersey, researchers first found that their reporting forms posed a number of problems, but a year after families began using redesigned forms, the number of filing problems to be handled by the experiment's office had dropped from 25 percent to 8 percent per reporting period.[8] On the basis of a series of comparisons between reports to the payments office and other sources of information in rural experiments, researchers concluded that although self-employed farmers did not perform as well as wage earners, "virtually all families were able to comply with the reporting requirements and to report information with a high degree of accuracy, lending strong support to the cost-saving administrative procedure of self-reporting by participants in welfare-type programs."[9]

Although Kenneth Kehrer believes that there was relatively little sustained interest in the monthly reporting and retrospective budgeting practices in Gary on the part of Indiana welfare officials, both he and Joseph Bell report that the innovations were influential in Colorado. Furthermore, as the experiments progressed, researchers began to ask if these techniques would not be profitably applied to the AFDC program. Based on results of her 1973 simulation study of the effects of a variety of retrospective budgeting approaches, Jodie Allen concluded that they could be.[10] Her work prompted the Department of Health, Education, and Welfare (and later, Health and Human Services) to undertake a series of demonstrations that studied the effects of a monthly reporting and retrospective budgeting program.

Findings from the first year of the earliest of these studies, conducted in Colorado, were instrumental in the enactment of the Reagan administration's 1981 budget proposal that mandated both practices nationwide for the AFDC and food stamp programs.[11] Of course, as is often the case when research results lead to a policy change, the findings were in accord with a prevailing sentiment, which had grown throughout the 1970s, that the welfare system should reduce its error rates and exercise firmer control over its disbursement of benefits.

Today, however, the verdict is still out on the ultimate usefulness of monthly reporting and retrospective budgeting. In contrast to the first-year findings on the Colorado program, results from the second year, which did not become widely available until after the 1981 federal mandate, showed that these innovations, rather than saving the state money, actually entailed a slight cost.

Apparently this reversal was a result of modifications in the conventional system to which the experimental practices were compared. During the first year of the study, AFDC redeterminations in the Colorado system were often conducted late and by mail, rather than face to face. In the second year, following a number of procedural reforms, Colorado had begun to institute timely face-to-face redeterminations.[12]

A broader summary of results from evaluations of monthly reporting and retrospective budgeting in five locations, including Colorado, is in keeping with these findings. The report, issued by Abt Associates Inc., concludes that monthly reporting produced savings larger than those under a conventional system that does not provide for face-to-face redeterminations but was no more effective than a system that uses such redeterminations.[13] Also, on the basis of data from the first year of the Colorado demonstration and from an Illinois demonstration, this same report found that "in practice, the differences between retrospective and prospective accounting in the demonstrations proved to be few and to affect a relatively small number of situations."[14]

What about the effects on recipients? Here, again, answers are mixed. Abt Associates concluded that neither monthly reporting nor retrospective budgeting caused particular problems for recipients. In Michigan and Massachusetts, the percentage of case closures due to failure to file monthly reports was similar to "analogous closures" for failure to appear for redetermination under the conventional system. Furthermore, interviews of people from closed cases in Illinois and Michigan, contacted from one to six months after closure, uncovered no significant differences in the value of AFDC benefits not paid to those who have been terminated under the monthly reporting and the conventional systems. Also in those states, a comparison of clients' circumstances in a given month to payments they received in the same month showed no significant differences in the extent to which the conventional and monthly reporting systems created lags in grant adjustment.[15]

Citing different data, the Center on Budget and Policy Priorities is less sanguine about the impact of monthly reporting on recipients. In a study of 883 AFDC households terminated in the Denver monthly reporting experiment in 1979 and late 1980 for failure to file a monthly report, Mathematica Policy Research estimated that between 20 and 50 percent of the households terminated for failure to file or failure to correct were actually eligible at the time of the termination. Another study conducted by the Michigan Department of Social Services found that more than nine out of ten recipients whose benefits had been terminated for failure to comply with the monthly reporting requirement were otherwise eligible for assistance when their cases were closed.[16]

Several observations can be made about these studies and the different directions in which they seem to point. John Bickerman, the co-author of the analysis of monthly reporting issued by the Center on Budget and Policy Priorities, points out that unlike the Mathematica and Michigan studies, the Abt research does not show how many of the cases under either system were closed in error. Also it does not show whether there was any difference between the number of families wrongfully terminated under each system. Bickerman also cites several factors that he believes would make monthly reporting more prone to error than less frequent face-to-face redetermination: the possibility of slip-ups in data processing, the potential problems associated with mailing forms, and the very frequency with which data are processed — the hypothesis being that information handled six times in six months rather than once every six months is more likely to be tainted with error just because it is transmitted more frequently.

In the face of the mixed evidence, it is somewhat difficult to know how to assess the retrospective requirement. In all likelihood, however, the presumption should be against heavy reliance on a system if there is a possibility that it can harm clients and little proof that it is superior to other methods. Concluding from the research findings that monthly reporting seems most useful when applied to cases where information would otherwise be missed, Bickerman and Greenstein recommend that states identify the categories of cases likely to be most troublesome to verify and target monthly reporting only to them.[17]

Interestingly, the Deficit Reduction Act of 1984 reversed in part the 1981 provisions on monthly reporting and retrospective budgeting, allowing states to cease these practices for all clients except those with earned income or a recent work history. This scaling down of the 1981 policy was apparently a response to a number of state officials who had objected to the changes.[18] However, according to the Office of Family Assistance, since the blanket mandate has been lifted only a handful of states have availed themselves of the opportunity to return to their previous methods of doing business. (The former Director of Governmental Affairs and Social Policy of the American Public Welfare Association believes that among states that have done so, concern for client welfare was probably the primary consideration.)

It is possible that most states have maintained monthly reporting and retrospective budgeting because, having become accustomed to these systems, they now have found that they can operate them fairly and that they do improve efficiency. In an informal discussion among administrators at a recent conference of the National Council of State Human Service Administrators of the American Public Welfare Association, a number of commissioners said they had found these systems useful.

It is also possible that some states have continued these practices simply because the momentum of the situation encourages them to do so; once the systems are in place, it may be easier to assume that they serve a useful function than it is to carry out a thorough review of what they do and do not accomplish. In any event, the persistence of unanswered questions about monthly reporting and retrospective budgeting over a decade after they were studied underscores the challenge of initiating and assessing even modest administrative changes in a large and complex income maintenance system.

While retrospective budgeting and monthly reporting were the administrative aspects of the negative income tax experiments with the most direct carryover into the welfare system, the experiments did raise other administrative issues that are of interest to welfare officials. One important question is the extent to which participating families were actually aware of the rules of the game. Periodic surveys of families participating in the rural experiments showed that rules concerning "the basic benefit level, the implicit tax rate, and the breakeven rate were understood by only about one-half of the families and that their understanding did not improve during the experiment. . . . More than one-quarter of the families thought that the program's tax on their income (which in fact ranged from 30 to 70 percent) was either zero or 100 percent."[19] Results of a survey administered to families after their first year in the Seattle-Denver income maintenance experiments showed that the mean percentage of correct answers for understanding the calculation of grant payments was 38 percent in Seattle and 46 percent in Denver.[20]

It is likely, as the analysts of this survey put it, that "average comprehension scores understate behaviorally related knowledge"[21] — in other words, that people are better able to act in accordance with rules than to answer questions about them on written tests. Nevertheless, these results raise a troubling question — namely, at what point policymakers defeat the purposes of the incentives they create by rendering them so complex that rewards and penalties are obscured. To cite just one contemporary example of this dilemma: How many AFDC recipients will choose to work or cease working on the basis of the rules for Medicaid eligibility for income earners that have been constructed and reconstructed over the past several years? A reasonable hypothesis would be that very few individuals have acted in terms of these exquisitely graded incentive structures.

Still another interesting question embedded in the negative income tax experiments is the extent to which an income maintenance system ought to be impersonal. In contrast to the welfare system, wrote David Kershaw and Jerilyn Fair, the New Jersey negative income tax experiment "was explicitly designed to minimize personal and face-to-face contact between participants and staff members."[22] Thus, when the New Jersey payments office ran into difficulties about a family's payment forms or benefits, field staff, who were responsible for relationships with the families, contacted them first with a form letter, then with a telephone call and next with a handwritten note. Home visits were a last resort.[23]

Although these practices represent a significant change from the model of heavy caseworker involvement in the affairs of a client family, the negative income tax experiments did not always succeed in making such a clean break with the conventional system. In the Gary experiments, for example, Kenneth Kehrer points out that for several reasons, many families actually had much more contact with professionals than was the case before they joined the experiment. First, some families in the experiment were offered support services — as families were in Seattle and Denver. Second, AFDC families in Gary were subject to the monthly reporting procedures of the experiment and to the usual AFDC six-month face-to-face recertifications. (Families on food stamps also underwent recertifications.) Finally, as in all the experiments, Gary families were interviewed by researchers, and often did not draw a distinction between the roles of these individuals and those of officials in an income maintenance system.

It is not even clear, moreover, that all of the experiments placed as high a priority on impersonality as New Jersey's did. Christopherson describes a system in Seattle and Denver whereby staff in local field offices contacted families about payment issues primarily by telephone, with occasional home visits, reserving mail only for routine information. As a result of contact, he reports, field workers often developed a family advocate role in disagreements with the payments department. Christopherson characterizes this relationship as a "productive" one, because field staff "understood the positions of both sides and could thereby best articulate the positions of each to the other."[24]

In all, this experience underscores an insight that often emerges from social policy experiments — that it is easier to decide in the abstract to keep a distance from participants and their concerns than it is to carry through on that resolution. Nonetheless, the question remains: However much the negative income tax experiments succeeded or did not succeed in developing a more impersonal income maintenance system, is such a goal a desirable one?

Adding further significance to this question is the way in which the effort to maintain impersonality is consistent with an important change in the AFDC program during the 1970s — the separation of income maintenance from social service functions. The driving force behind this reform was philosophic but budgetary aspects were important: by assigning income maintenance functions to workers who commanded relatively low wages, the program could concentrate limited dollars for social services on salaries for a smaller number of professionals. It was argued that the separation would allow social workers to devote their full energies to assisting clients. Still, an important point made in debates over this modification echoed the thesis of the negative income tax — that income maintenance should be a less meddlesome, client-involved system, that the bureaucrat who decides what benefits are due a client should not be peering over that person's shoulder.

In many ways, this is an appealing argument, and in one respect it seems even more cogent today than it was when the income maintenance experiments were first designed. Since that time, we have acquired a clearer vision of the nature of the welfare caseload. The recent work of Mary Jo Bane and David Ellwood on caseload dynamics dispels the notion of the typical welfare recipient as a person sunk into dependency.[25] Most recipients, we know, leave welfare rather rapidly — and in all probability many in this group do not require massive or intensive special assistance delivered through the welfare system. Temporary income support and modest help with finding child care or with conducting a job search may be all that is required.

But Bane and Ellwood also alert us to the existence of a significant number of people in the caseloads — the long-term recipients who consume a disproportionate share of the nation's welfare dollars. Six out of ten people who do not leave the rolls at the end of two years, their study shows, are likely to be there at the end of six.

These chronic recipients are likely to need more intensive social services than many short-stayers. And when the system does encounter this level of need, it seems dysfunctional to assign clients with multiple disadvantages to an income maintenance worker who knows nothing about the complex set of problems that may well be contributing to their dependency. What seems to make more sense is a case management approach, with one professional asked to become familiar with all the relevant circumstances, both financial and social service, that pertain to a person's stay on welfare. In this situation, the effort not to become intrusive is considerably less compelling than the need to deliver services in a coherent manner.

Conclusions

Today, as reformers of all political persuasions appear poised to make another attempt to improve a welfare system that pleases almost no one, can they look to the negative income tax experience for guidance? In one respect, they should do so. Especially for the many recipients who use welfare as a temporary source of aid, the goal of a simplified approach to income maintenance merits serious attention. At the same time, the design of the experiments may serve to remind contemporary reformers that it is easier to espouse such an ideal than it is to translate it into policy.

The overall thrust of the negative income tax effort may be less relevant to the plight of chronic recipients. While two of the experiments did make a bow in the direction of support services, this was not their main focus. Yet today, it appears that a priority for AFDC recipients with multiple and long-term disadvantages is not only to offer them income maintenance but also to ensure that the welfare system identifies them as early as possible and then helps them attain access to a coordinated and sustained array of services.

Meanwhile, as the first section of this paper has suggested, a review of the negative income tax experiments as a particular style of evaluation with a particular approach to reform suggests that as policymakers embark on new research efforts, they should try to trace out how their investigations will relate to the systems they are intended to improve. As noted previously, experiments cannot and should not be all cast in one mold. However, in all cases it seems worthwhile for researchers to devote attention to the problem of how, in view of the predictable constraints and gaps between the worlds of theoretical research and functioning political systems, understanding and communication can be maximized. Simplicity of research design and sharp framing of questions are important, as is the need to provide for good cross-site analysis. Every effort should be made to convey research information clearly and simply, and to disseminate it periodically, rather than only at the end of the research study. With all of this, it remains a challenge to design and conduct theoretical research that matters in the world of public policy. Reviewing the negative income tax experiments has offered an opportunity to assess the dimensions of this challenge.

[1]Mitchell Ginsberg and Norman Lourie, "The Current Status of Human Services," *Public Welfare,* Summer 1974, p. 29.

[2]See David Kershaw and Jerilyn Fair, *The New Jersey Income Maintenance Experiment, Volume 1, Operations, Surveys and Administration.* New York: Academic Press, 1976, Chapter 12.

[3]Information from Kenneth Kehrer in this article is taken from telephone conversations with the author.

[4]See Board of Directors, Manpower Demonstration Research Corporation, *Summary and Findings of the National Supported Work Demonstration.* Cambridge, Mass.: Ballinger Publishing Company, 1980.

[5]See Judith M. Gueron, *Work Initiatives for Welfare Recipients.* New York: MDRC, 1986.

[6]Gary Burtless, "The Work Response to a Guaranteed Income: A Survey of Experimental Evidence," paper prepared for this conference.

[7]Eric A. Hanushek, "Non-Labor-Supply Responses to the Income Maintenance Experiments," paper prepared for this conference.

[8]Kershaw and Fair, *The New Jersey Income Maintenance Experiments, Volume I,* p. 67

[9]D. Lee Bawden and William S. Harrar, "Design and Operation," in John L. Palmer and Joseph A. Pechman, editors, *Welfare in Rural Areas: The North Carolina-Iowa Income Maintenance Experiment.* Washington: The Brookings Institution, 1978, pp. 39, 41.

[10]See Jodie T. Allen, *Designing Income Maintenance Systems: The Income Accounting Problem.* Washington: The Urban Institute, 1973.

[11]John Bickerman and Robert Greenstein, *Research Findings on Monthly Reporting Systems and Their Implications for State Administrators.* Washington: Center on Budget and Policy Priorities, November 1983, p. 8.

[12]John A. Burghardt, *Impact of Monthly Retrospective Reporting Requirements: Evidence from the Second Year of the Colorado Monthly Reporting Experiment,* Princeton: Mathematica Policy Research, Inc., February 1982, cited in Bickerman and Greenstein, *Research Findings,* pp. 9-10.

[13]William L. Hamilton, *Monthly Reporting in the AFDC Program: Executive Summary of Demonstration Results.* Cambridge, Mass.: Abt Associates Inc., September 1985, p. iv.

[14]Ibid., p. 20.

[15]Ibid., p. 35.

[16]David A. Price, *Study of AFDC Cases Discontinued by the Colorado Monthly Reporting System,* Princeton: Mathematica Policy Research, March 1981, cited in Bickerman and Greenstein, *Research Findings,* pp. 18-20. Department of Social Services, State of Michigan, *The Impact of Monthly Reporting in Michigan,* Volume I, September 1983, p. IV-10.

[17]Bickerman and Greenstein, *Research Findings,* pp. 27-31

[18]See Children's Defense Fund, "Other Rule Changes Concerning Payment Amounts," *CDF Reports,* special issue on the Deficit Reduction Act of 1984, August 1984.

[19]Bawden and Harrar, "Design and Operation," p. 39.

[20]Gary Christophersen, *Final Report of the Seattle-Denver Income Maintenance Experiment, Volume 2: Administration.* Princeton: Mathematica Policy Research, p. 101.

[21]Harlan Halsey, Binn Muraka and Robert Spiegelman, "The Study of Participant Comprehension of the Seattle and Denver Income Maintenance Program," Menlo Park, Ca: SRI International, 1979, quoted in Christophersen, *Final Report,* p. 101.

[22]Kershaw and Fair, *The New Jersey Income Maintenance Experiments,* p. 67.

[23]Ibid., p. 69.

[24]Christophersen, *Final Report,* pp. 97-8.

[25]See Mary Jo Bane and David Ellwood, *The Dynamics of Dependence: The Routes to Self-Sufficiency,* Cambridge, Mass.: Urban Systems Research and Engineering, 1983.

Discussion

*Wilbur J. Cohen**

Barbara Blum's paper is excellent and clear. It credits the negative income tax experiments with three specific administrative accounting changes in the aid to families with dependent children (AFDC) and food stamp welfare programs: monthly reporting, automatic processing, and retrospective budgeting. She points out, however, that "the verdict is still out on the ultimate usefulness" of these changes mandated by the Reagan administration's 1981 legislation. The negative income tax experiments, therefore, cannot be said to have produced no results so far. Social science research must always be thankful for even small results in the short run. The results in the long run probably will depend on how long the research community is willing to wait. I believe there will be other results. The ultimate lessons of the experimentation, in my opinion, are likely to result in incremental improvements in our welfare programs over time rather than a single "quantum leap."

Ms. Blum's paper indicates that the administrators of the negative income tax experiments did not have a close working relationship with state welfare administrators. I do not believe this was a fatal mistake. Nevertheless, the results of the experiments have not been widely explained to state welfare personnel, and thus these results have not been shared among those who might be able to put them to administrative use or at least discuss them with their governors. But here again, I do not believe this would have made any more difference than having discussed the effect of the marital dissolution results with Senator Moynihan. Despite the original endorsement of the negative income tax

*Professor of Public Affairs, Lyndon B. Johnson School of Public Affairs, University of Texas at Austin.

by Milton Friedman in 1962, the basic idea has had no favorable impact on the Reagan administration.

Ms. Blum's paper contrasts these income maintenance experiments and the Manpower Demonstration Research Corporation's experimentation on "workfare," the Work Incentive (WIN) Program, or related work and training projects. MDRC utilized welfare clients as the subjects of its research, which involved a close relationship with state welfare administrators. It has taken 25 years to develop work and training experiments with welfare clients since the idea was first put forward in 1961-62 in the Community Work and Training Program. The results from this research are not entirely favorable, but at least they have not produced the negative results of the income maintenance experiments with regard to reduction in hours worked, or suggested that such payments contribute to marital dissolution. I venture to prophesy that work and training experimentation will spread in the next decade; economists should spend more of their energy in the design and evaluation of the projects.

One area that seems to me to call out for research experimentation is the appropriate "earnings disregard" for work income. The current AFDC program's $30 and one-third earnings disregard, which I negotiated in 1967 (and which took me nearly seven years to obtain), has been around for nearly 20 years and needs reexamination. Originally I proposed $50 and 50 percent to the House Committee on Ways and Means in 1967, then offered them $40 and 40 percent as a compromise. They took $30 and one-third, on the grounds that further testing of the appropriate formula was needed in terms of actual experience. Yet, as far as I know, no extensive research has been done on this matter. Why? The earnings disregard is such a fundamental element of any income maintenance program. I would like to see the federal AFDC law give states wide latitude to experiment with different earnings disregards.

A major difficulty with any income maintenance proposal is the level of payment: if the proposal sets too high a payment for the person with zero income and includes too high an earnings disregard, many persons will receive payments at levels far above what others believe are desirable or even financially feasible. It does not seem possible to make provision for abolishing poverty as an income strategy and at the same time include a significant income incentive for those persons who can or want to work.

The fact of the matter is that a negative income tax consists of several major elements:

1. A basic floor of income for persons with no earnings;
2. An earnings disregard, as an incentive to work;
3. Disregard of assets as a bar to receipt of payment;

4. Simplified administration by nonprofessional personnel; and
5. A presumption that the plan will be administered on a nationwide basis by a non-welfare agency.

In my opinion, it would be possible to have different states put such a plan into operation with different income, asset, and earnings disregards. Senator Ribicoff advocated several such state experiments in 1970 at the time the Family Assistance Plan proposal failed in the Senate. Until now no state, the federal government, or any economist has picked up this challenge, however.

I think we should experiment with this and some other policies that have been advocated for some time but have not been put into operation. They include:

1. Federal legislation requiring a program, in all states, which would provide assistance payments under AFDC to the children of needy unemployed parents, allowing states a wide latitude on the definition of who is "unemployed," but with some minimum federal definition of the term "unemployment."
2. A minimum federal floor of income support, set initially at about 65 percent of the poverty threshold with both AFDC income and food stamp evaluation to be counted. This floor would rise 2 percent each year for 10 years, to 85 percent of the poverty threshold.
3. A large-scale work and training program for all welfare clients who want to participate.
4. State experiments with earnings disregards.
5. A broadened Medicaid program that would include all needy individuals with incomes below the poverty line and thus divorce income determination for Medicaid from the AFDC program.
6. Federal and state funds for evaluation of these policies.
7. A federal advisory council to report to Congress periodically on the policies and research and to make recommendations on them.

I believe we could have these policies in full effect by 1995. We could then take another look at where we could and should go from there.

Lessons for Future Public Policy and Research

*Richard P. Nathan**

Social experimentation with random assignment is unique to the United States as an approach to public policy research. As far as I can determine, no other Western countries have conducted social experiments with a randomly assigned group. In the United States, we have now had sufficient experience with this form of policy research that it is appropriate to take a hard look at its conduct and its usefulness in the governmental process.

This paper examines the income maintenance experiments and their implications for social policy and for future research. Each topic is introduced by a section on their lessons for the history of applied social science. To organize the material I have adopted the approach recommended by Richard E. Neustadt and Ernest R. May.[1] They urge policy analysts to draw on the historical record to answer three questions: What is known? What is unclear? What is presumed?

Before proceeding, it is necessary to add a comment about the point of view I bring to this analysis. As a former government official involved in the development of Nixon's Family Assistance Plan, I have come to the conclusion that the Plan was a mistake. I believe the mistakes made in developing and advocating the Family Assistance Plan were a result of its heavy dependence on the idea of the negative income tax, which was riding high in the public policy research and analysis community at the time Nixon's plan originated.

As regards the second main topic considered in this paper — the research implications of the negative income tax experiments — I have a

*Professor of Public and International Affairs, Woodrow Wilson School, Princeton University.

bias in favor of large, systematic demonstration research projects to test possible new policies, conducted on a basis that involves the random assignment of participants to treatment and control groups. However, I have considerable reservations about the negative income tax experiments. I believe demonstrations of new *service-type* policies are more likely to produce useful results for government policymakers than demonstrations to test universal *income-transfer* initiatives.

Lessons and Implications for Social Policy

When a new policy is tested, ideally we would like to know about its macroeconomic effects and its microeconomic effects. But economic effects are not the whole story. A public policy change like the negative income tax would be expected to have social and psychological effects as well. For recipients, it could increase or diminish self-esteem, work or school motivation, health, and happiness. Likewise, it could affect the happiness — call it a feeling of altruism or a sense of security — of a community, a neighborhood, or the society as a whole. In the past, these social and psychological dimensions of the effects of policy change often have been left out in the conceptualization and planning of social experiments. In addition, we are likely to be interested in the political and institutional effects of a potential new policy being tested in a social experiment. How would a negative income tax affect the managerial capacity, finances, and relative roles of different levels of government and different types of public and private agencies and organizations?

The negative income tax experiments were launched in the late sixties at a time when economics was riding high as the lead discipline for applied social science in government in the United States. As a result, it is not surprising that economic effects, specifically labor market effects, were highlighted in the design and execution of the experiments.[2] This emphasis is reflected in the discussion which follows, using the Neustadt-May approach in examining the history of what is known, unclear, and presumed from the experiments.

What Is Known?

When the debate about the Family Assistance Plan was at its peak in the U.S. Senate in 1970, officials in the Nixon administration, in a move that is still debated, issued a "preliminary" report on the results of the New Jersey negative income tax experiments. This report described the effects of the plan tested in New Jersey on work incentives for adults in two-parent low-income families. This issue of the work-incentive effect of subsidies to working-age, able-bodied adults has deep roots in social policy in the United States. Observing the British Poor Laws in 1766, Benjamin Franklin summed up his reaction: "In short, you offered a premium for the encouragement of idleness, and you should not now wonder that it has had its effect in the increase of poverty."[3]

The conclusion of the 1970 preliminary report on the results of the New Jersey experiments was that Benjamin Franklin was wrong. There was no work-disincentive effect in the early New Jersey returns. Senator John Williams, Republican of Delaware, at this juncture called on the General Accounting Office to "audit" these results. Its finding was that this report by the Office of Economic Opportunity was "premature," which clearly it was.[4] Later findings showed that there was a labor-supply cost to the experiments. This work-disincentive effect was found to be greatest for women heading single-parent welfare families in the Seattle-Denver experiments, although it also showed up for men in two-parent families in both the New Jersey and Seattle-Denver experiments. Gary Burtless and Robert H. Haveman concluded: "The Seattle-Denver experiment has played a useful role in overturning the notion, especially popular among economists and idealistic reformers, that lower marginal tax rates are automatically associated with a greater stimulus to work."[5]

The Seattle-Denver experiments also showed that the negative income tax plan they encompassed tended to have an adverse effect on family formation and to encourage family breakup.[6] These findings, along with the labor market findings and their cost implications discussed below, weakened the case for this type of welfare reform, especially during the Carter administration.

What Is Unclear?

Even if we know about the labor market effects of a negative income tax, this does not necessarily mean that we can generalize about these effects as they apply to a new and widely publicized national scheme. The reason for this is that we do not know how the adoption of a "guaranteed-income" program would be interpreted by the eligible population. This is the issue of external validity. We are interested in this connection in how such a universal, broad-gauged policy change would affect the tastes of society as a whole. It is in these terms that the Coyle-Wildavsky paper for this conference discusses the cultural dimension of policy change. Would a negative income tax be seen as governmental support for added leisure?[7] Such an outcome could accentuate the labor disincentive effect, which as just noted was found to have occurred in the New Jersey and Seattle-Denver experiments.

At an April 1974 Brookings conference on the New Jersey experiments, Peter H. Rossi presented a paper on the non-labor-force responses to the experiments.[8] Rossi said it was paradoxical that, despite the heavy reliance of sociologists on the collection of primary data and the extensive use by psychologists of experimental designs, it was economists who "played the major role in designing and fielding the income maintenance experiment."[9] The result, said Rossi, is that we know very little from the experiments about the noneconomic effects of a negative income tax on individuals, although these effects were often presumed by the sponsors of the research.

On a general basis, Thomas Pettigrew has expressed concern about the neglect of the psychological dimension in poverty research.[10] He has commented, for example, on the psychological concept of "learned helplessness as an unintended, but important, possible consequence of transfer payments." A similar point is reflected in the current social policy criticism of conservatives like Charles Murray, who writes about how the dependency effects of welfare programs have undermined the self-image of recipients.[11]

What Is Presumed?

The most important presumption (for purposes of this paper) made by the sponsors of the experiments is reflected in the point made by Rossi, that the sponsors assumed that if the negative income tax "worked" in the labor market, the case for it on other grounds was a strong one. In doing this, the sponsors of the experiments failed to take adequate account of fundamental political and strategic issues related to the idea of a negative income tax. I discuss these issues in the section that follows.

Implications for Social Policy

The main effect of the income maintenance experiments was to educate government participants as well as the media and interested citizens on the policy issues raised by the idea of a negative income tax. Controversies arose in the Nixon period about the work-incentive effects of the experiments, and in the Carter period about their cost and their effect on family breakup. The educational process that ensued was expensive and sowed seeds of doubt about an idea that originally had been seen as a bold solution for many of the nation's social ills.

The economic conundrum of the negative income tax is demonstrated if we set the income guarantee at an "adequate" or "near-adequate" level for families and then want to have a marginal reduction rate in the 50 to 60 percent range. The added cost of such a plan, if its coverage is comprehensive and its rate structure a smooth curve (that is, un-notched), is bound to be very large — on the order of $20 billion to $25 billion. Unless the society had been ready to devote such large additional amounts of money to a redistributive scheme for the working-age, able-bodied poor, the negative income tax approach was not in 1969 (and is not today) a feasible national policy option.

Presidents Nixon and Carter both gave evidence of understanding this issue, at least intuitively. From the outset, Nixon stressed that the Family Assistance Plan was not a guaranteed income. The work requirement was to be serious and enforced. His plan also included a substantial amount of money for public service jobs for eligible family heads for whom suitable jobs were not available elsewhere. President Carter went even further in his welfare reform scheme, which like Nixon's embodied negative income tax features. He advocated a guaranteed (or close to guaranteed) job for eligible heads of welfare families for whom regular employment was not available.

The essence of the Nixon position, the Carter position, and even more so the Reagan position, is: *Money alone is not the answer.* The negative income tax approach, grounded in neoclassical economics, was never a comfortable one for most politicians.

One wonders in this context why the experiments were undertaken. Were they a delaying tactic supported by political officials who resisted this approach to welfare reform? Or were they an effort on the part of proponents of a negative income tax to put their idea on the agenda and prove that it would work? I think it was the latter. If I am right, then the issues are: Who should set the agenda for policy research? And, how should this be done? Experiments are expensive both in budgetary terms and in terms of their opportunity cost for the social science policy research community. *I believe social experiments should be restricted to situations in which three conditions apply. Politicians need to be: (1) genuinely interested in dealing with an issue; (2) uncertain about how to do so; and (3) willing to consider the approach that is the subject of experimentation.* In my view, the negative income tax experiments did not satisfy these conditions. There are ways, as discussed below, in which these conditions can be satisfied by social experiments in the welfare policy field.

To summarize, three strands of opinion have dominated the welfare reform debates over the past 20 years. One is the income strategy favored by the policy analyst and a few liberal politicians.[12] A second strand in the welfare reform debate is the block grant or devolutionary position favored by conservatives, with Ronald Reagan in the forefront both as a governor of California and as President. The third position, the employment approach to welfare reform, has both liberal and conservative adherents. A jobs component was featured in both the Nixon and Carter welfare reform plans. The employment approach is also central to the so-called "workfare" component of Reagan's approach to welfare reform.

As the negative income tax or income strategy lost its lustre, new departures on the employment front have increasingly dominated policy debates. Under Reagan, employment approaches to welfare reform are now the subject of new social experiments in many states. We need to step back and look at the background of Reagan's position in order to consider these new work/welfare experiments.

Since the early seventies, the Reagan approach to welfare reform has specifically rejected the negative income tax concept of setting the marginal reduction rate for welfare benefits at a level that will stimulate work effort. In 1981 Reagan was successful in reducing the effect of the $30 plus one-third deduction for aid to families with dependent children (AFDC), first enacted in 1967.[13] In the Omnibus Budget Reconciliation Act of 1981, Reagan also tried to require states to assign all eligible AFDC family heads (employable adult recipients with children over six years of age) to ''workfare'' jobs, provided child-care was available. Although Congress would not go along with a compulsory and universal ''workfare'' requirement, provisions were included in the 1981 budget act allowing the states to test the work-for-your-welfare approach as well as other employment approaches to reducing welfare dependency. Over two-thirds of the states are now taking advantage of this new authority under the heading of ''workfare.'' This term has come to have a broader and more liberal meaning in the 1980s, applying not just to the work-for-your-welfare idea, but also to new approaches that require employable heads of welfare families to participate in job placement, training, and educational services, as well as community employment programs.

The Manpower Demonstration Research Corporation, based in New York City, is conducting eight state-based demonstration studies with random assignment on variations of the ''workfare'' approach. Altogether, more than 35,000 people have been assigned either to a control group or to new programs for job counseling, job preparation, and community work experience. The results of these demonstrations so far have been promising, although the earnings and work increases achieved are not all that large and there is variation among the states in these terms.[14] In effect, the states are serving as testing grounds for the employment approach to welfare reform, on a basis that I believe satisfies the three conditions described above and that involves a delicate political balancing act by liberals and conservatives. The question raised by these state workfare initiatives is whether a skillful blend of new employment-oriented program features and procedures can avoid the dilemma of welfare reform described by Henry Aaron in 1984. Looking back at the work requirement of the welfare reform debates of the seventies, he wrote:

> The acknowledgement of the need for a work requirement created an insoluble dilemma, however. With a sufficiently coercive administrative system, potential welfare recipients could be required to accept existing low-quality, low-wage jobs in the private sector. If enough private sector jobs did not exist, public sector positions could be created at low cost. If a work requirement discouraged enough people from applying for welfare, costs might even be reduced. But the coercion that would be necessary to make such a requirement work violated notions of fairness and rights. Alternatively, the public sector could create jobs with sufficiently attractive working conditions and wages to reduce greatly the need for coercion. But the size of the program would be unprecedented, and its cost would be prohibitive, particularly since many workers in unattractive private sector jobs would find it expedient to switch to superior public sector jobs. Trapped on this political Moebius strip, welfare reform went nowhere.[15]

Although all the results are not in yet, the work/welfare experiments show promise of offering a way out of the trap described by Aaron on a basis that the public and politicians would be likely to favor.

Lessons and Implications for Policy Research

What Is Known?

Bette and Michael Mahoney described the New Jersey negative income tax experiments as "an experiment in experimentation."[16] To the credit of the researchers involved, we learned from those experiments that it is possible in the United States to conduct large-scale, rigorous, honest demonstration research projects with the random assignment of participants to treatment and control groups. We also learned that such research is expensive and that it takes a long time to conduct. The New Jersey experiment was first proposed by the Office of Economic Opportunity in 1967. It was conducted from 1968 to 1972. Reports were issued late in 1973 and in 1974, by which time Nixon's Family Assistance Plan had already been abandoned. The Seattle-Denver experiments were started in 1970 and involved over three times as many participants as the New Jersey experiments. They took over a decade to complete.

Both experiments can be said to have worked in research terms. This is a real achievement when one considers the hurdles that must be cleared in mounting and conducting social experiments in complex real-world settings: selection bias, contamination, obtaining informed consent, establishing good working relations with program operators, collecting full and accurate data, locating respondents (especially controls), and avoiding sample attrition are examples.

What Is Unclear?

Despite the fact that the experiments worked from a technical point of view, it is not clear that they achieved their purpose as an input to government policymaking. Actually, this question has two parts: (1) Did politicians (broadly defined to include elected and appointed officials and the representatives of major organizations) use the results of this demonstration research? and (2) Did the results of the research achieve the aim of the sponsors, which was to advance the idea of a negative income tax?

One interpretation is that the results of the experiments were used, but not in the way the sponsors of the research intended. I believe that the very fact that the experiments were being planned and conducted had an educational effect in the early Nixon years and on balance advanced the concept of a negative income tax (although mistakenly). However, later on in the Carter period the results of the experiments tended to undermine the negative income tax idea. These negative political results of the experiments — that is, negative for a negative income tax — may have destroyed the chances of enacting comprehensive welfare reform with a welfare reduction rate set at a level that would not undermine the work incentives of recipients.

What Was Presumed?

As already mentioned, the historical record suggests that the sponsors of the experiments saw them as a way to dispel the belief that a negative income tax ''would bring about a large increase in idleness among those who would otherwise have worked.''[17] Heather Ross, working first at the Council of Economic Advisers and later at Health, Education, and Welfare and at Brookings,[18] developed the original plan for the demonstrations. However, even in this period, there was considerable uneasiness about these experiments. When the time came to announce the start of the experiments, officials (particularly Sargent Shriver as head of the Office of Economic Opportunity) had second thoughts. Shriver decided to proceed, but to do so on a low-key basis.[19] The initial contract for the New Jersey experiment was financed using previously appropriated research funds, and its announcement was withheld until after Congress had recessed for Labor Day in 1967.

In my view, it is perfectly appropriate for the sponsors of social science research and for the lead researchers to believe in the programs they are testing. This is actually easier if their research involves random assignment, which has the effect of insulating demonstration researchers against manipulation. Random assignment keeps us honest, and in this way reduces the problem of bias toward the tested plan on the part of both research sponsors and researchers. The problem was not that the negative income tax experiments, as defined in a limited way, were poorly done, although there were important difficulties along the way. The problem lies in an often neglected area for social policy writing, involving the connection between the research process and the political process. The final section of this paper addresses the implications of the negative income tax experiments for public policy research.

Implications for Policy Research

The fact that the findings of the income maintenance experiments undercut rather than supported the case for a negative income tax is not a bad outcome. While it is no doubt of little comfort to the originators of these experiments, this outcome speaks well for the integrity of large-scale social experiments. Nevertheless, I believe that on balance these particular experiments were not well advised. The main reason for this conclusion involves the conundrum of the negative income tax. The cost of covering millions of additional people under an income maintenance scheme presented a policy choice that simply was not in the cards in the latter part of the 1960s, even if the labor-supply analysis from the experiments had revealed no adverse or even a positive work-incentive effect.

There was also the related psychological dimension or stigma of adding large numbers of new people to the welfare rolls; this subject was not addressed in the design and conduct of the experiments. Taking into account their underlying political values and aims, I believe the sponsors of the negative income tax experiments allowed their research agenda to get ahead of their political agenda. Unless large amounts of additional money could have been obtained for a new and expanded welfare reform system, we were — and still are — better served by a multi-track welfare system. One track — AFDC — in many states has a near-''adequate'' benefit and a high welfare reduction rate to serve non-working family heads. A second track that particularly aids the working poor is the food stamp program, which in essence is a mini negative income tax that has a relatively low reduction rate and is now almost fully fungible. Various supplements (school lunches, Medicaid, housing subsidies) augment this assistance in ways that expand the political constituency for aiding the nation's most controversial dependent population.

In my view, such a multi-track system of different strokes for different folks is intellectually preferable to a negative income tax, absent the willingness on the part of the society to make a major new resource commitment (in the range of $20 billion to $25 billion) to income redistribution to the poor. Farmers, grocers, doctors, hospitals, builders, realtors, and educators (for school lunches, compensatory education, and college scholarships, for example) all have a stake in helping the poor under current conditions — or at least they perceive that they do.

In this context, one function of the negative income tax experiments was to teach these lessons. Unfortunately, however, this is not a cost-effective teaching strategy. I conclude that the experiments were unwise, but that the idea of social experiments with random assignment which they introduced is a good one. In particular, I believe one of the main implications of the negative income tax experiments for policy research is that experiments of more selective service-type policy initiatives are to be preferred over demonstrations of universal income-transfer schemes (for example, cash, health insurance, housing). The problems with testing universal income-transfer schemes are twofold: (1) the underlying value issues are bigger and more difficult to deal with than in the case of tests of variants of social service programs; and (2) the effect of a new policy on behavior is more likely to be pervasive and important in the case of a universal and highly visible new program like a negative income tax than in the case of service-type programs. Subsequent experiments in the field of social policy, notably those conducted by the Manpower Demonstration Research Corporation as described in the paper for this conference by Barbara Blum, have benefited greatly from the earlier experience of the negative income tax experiments.

Another critical research implication of the experiments, brought out in papers and discussions by Lee Rainwater and others at this conference, is that once we decide to embark on a social experiment we should *seek to learn more* from such endeavors than we did in this case. The dominance of economists in the negative income tax experiment had important consequences. In leaving out sociologists, psychologists, and political scientists as major players, the sponsors of the experiments, in effect, left out variables from the research equation that are important both to politicians and to society as a whole.

In sum, the fact that the negative income tax experiments worked is important for the future, despite my conclusion that the subject for experimentation was in this case not well advised. The planning of demonstration research involves both art and science. The negative income tax experiments, as the first such effort of this type, led the way in developing both the capacity and the sensitivity necessary to the more effective use of social experimentation as an input to the governmental process.

[1]Richard E. Neustadt and Ernest R. May, *Thinking in Time: The Uses of History for Decision Makers* (New York: The Free Press, 1986).

[2]Vincent J. Burke and Vee Burke, *Nixon's Good Deed, Welfare Reform* (New York: Columbia University Press, 1974). See especially pp. 14-24.

[3]As quoted in Gertrude Himmelfarb, *The Idea of Poverty* (London, Boston: Faber and Faber, 1984), p. 5.

[4]Daniel P. Moynihan, *The Politics of a Guaranteed Income: The Nixon Administration and the Family Assistance Plan* (New York: Random House, 1973) pp. 509-512.

[5]Gary Burtless and Robert H. Haveman, "Policy Lessons from Three Labor Market Experiments," in R. Thayne Robson, ed., *Employment and Training R & D, Lessons Learned and Future Directions* (Kalamazoo, Mich.: W.E. Upjohn Institute for Employment Research, 1984) p. 111. See also Burtless's paper for this conference.

[6]The paper by Glen G. Cain in this volume suggests that these effects may not be as strong as was first suggested.

[7]"Comment" by Richard P. Nathan, in Joseph A. Pechman and P. Michael Timpane, eds., *Work Incentives and Income Guarantees: The New Jersey Negative Income Tax Experiment* (Washington: The Brookings Institution, 1975), p. 199.

[8]Peter H. Rossi, "A Critical Review of the Analysis of Nonlabor Force Responses," in Pechman and Timpane, eds., *Work Incentives and Income Guarantees*, pp. 157-182.

[9]Ibid., p. 161.

[10]Thomas F. Pettigrew, "Social Psychology's Potential Contributions to an Understanding of Poverty," in Vincent T. Covello, ed., *Poverty and Public Policy: An Evaluation of Social Science Research* (Boston, Mass.: G.K. Hall, 1980).

[11]Charles Murray, *Losing Ground: American Social Policy, 1950-1980* (New York: Basic Books, 1984).

[12]In the latter part of the seventies, some conservative politicians, notably Caspar Weinberger following Milton Friedman, advocated the negative income tax approach, basically as a way to consolidate and control welfare spending.

[13]Two years later, Congress diluted this Reagan victory by continuing the $30 deduction, but not the one-third deduction, beyond four months.

[14]Judith M. Gueron, *Work Initiatives for Welfare Recipients* (New York: Manpower Demonstration Research Corporation, March 1986).

[15]Henry J. Aaron, "Six Welfare Questions Still Searching for Answers," *The Brookings Review*, Fall 1984, p. 14.

[16]Bette S. Mahoney and W. Michael Mahoney, "Policy Implications: A Skeptical View," in Pechman and Timpane, eds., *Work Incentives and Income Guarantees*, p. 197.

[17]Ibid.

[18]Robert A. Levine, "How and Why the Experiments Came About," in Pechman and Timpane, p. 17.

[19]Ibid.

Participants

HENRY J. AARON, *Senior Fellow,* The Brookings Institution
ORLEY C. ASHENFELTER, *Professor of Economics* and *Director,* Industrial Relations Section, Princeton University
BLANCHE BERNSTEIN, *Consultant, Social Welfare Policy*
BARBARA B. BLUM, *President,* Foundation for Child Development, and former *President,* Manpower Demonstration Research Corporation
KATHARINE L. BRADBURY, *Economist,* Federal Reserve Bank of Boston
GARY BURTLESS, *Senior Fellow,* The Brookings Institution
GLEN G. CAIN, *Professor of Economics,* University of Wisconsin
WILBUR J. COHEN, *Professor of Public Affairs,* Lyndon B. Johnson School of Public Affairs, University of Texas at Austin
JOSEPH P. CORBETT, *Chief of Planning, Policy Control and Review Group,* U.S. Department of Labor
DENNIS J. COYLE, *Doctoral candidate* and *Research Assistant,* Survey Research Center, University of California at Berkeley
SHELDON DANZIGER, *Professor* and *Director,* Institute for Research on Poverty, University of Wisconsin at Madison
DAVID T. ELLWOOD, *Associate Professor of Public Policy,* John F. Kennedy School of Government, Harvard University
RICHARD F. ELMORE, *Professor of Education,* Michigan State University, and *Senior Research Fellow,* Center for Policy Research in Education.
EDWARD M. GRAMLICH, *Acting Director,* Congressional Budget Office, and *Professor of Economics and Public Policy,* University of Michigan
DAVID Z. GREENBERG, *Professor of Economics,* University of Maryland, Baltimore County Campus
JUDITH M. GUERON, *Executive Vice President,* Manpower Demonstration Research Corporation
ROBERT E. HALL, *Professor of Economics* and *Senior Fellow,* Hoover Institution, Stanford University
ERIC A. HANUSHEK, *Professor of Economics,* University of Rochester
JERRY A. HAUSMAN, *Professor of Economics,* Massachusetts Institute of Technology
LEONARD J. HAUSMAN, *Associate Professor of Economic and Social Policy,* Heller School, Brandeis University
ROBERT H. HAVEMAN, *Professor of Economics,* University of Wisconsin
HUGH HECLO, *University Professor,* George Mason University; formerly *Professor of Government,* Harvard University
ROBINSON G. HOLLISTER, JR., *Professor of Economics,* Swarthmore College
MICHAEL C. KEELEY, *Senior Economist,* Federal Reserve Bank of San Francisco
ROBERT I. LERMAN, *Director of Research,* Center for Human Resources, Heller School, Brandeis University
ROBERT A. LEVINE, *President,* Canyon Analysts
FRANK LEVY, *Professor of Public Affairs,* School of Public Affairs, University of Maryland
DAVID LINDEMAN, *Principal Analyst,* Congressional Budget Office
IRENE LURIE, *Associate Professor of Public Administration and Public Policy,* State University of New York at Albany
JOSEPH L. McGAVICK, *Partner,* Deloitte Haskins & Sells
LAWRENCE M. MEAD, *Associate Professor of Politics,* New York University
CHARLES E. METCALF, *President,* Mathematica Policy Research, Inc.

ROBERT T. MICHAEL, *Director*, National Opinion Research Center and *Professor of Education*, University of Chicago
ROBERT A. MOFFITT, *Professor of Economics*, Brown University
WILLIAM A. MORRILL, *President*, Mathtech, Inc.
FRANK E. MORRIS, *President and Chief Executive Officer*, Federal Reserve Bank of Boston
ALICIA H. MUNNELL, *Senior Vice President and Director of Research*, Federal Reserve Bank of Boston
CHARLES MURRAY, *Bradley Fellow*, Manhattan Institute for Policy Research
RICHARD P. NATHAN, *Professor of Public and International Affairs*, Woodrow Wilson School, Princeton University
GUY H. ORCUTT, *Professor of Economics and Statistics*, Institution for Social and Policy Studies, Yale University
LARRY L. ORR, *Director*, Abt Associates
JOHN L. PALMER, *Senior Research Associate*, The Urban Institute
JOSEPH A. PECHMAN, *Senior Fellow*, The Brookings Institution
LEE RAINWATER, *Professor of Sociology*, Harvard University
ROBERT D. REISCHAUER, *Senior Fellow*, The Brookings Institution
PHILIP K. ROBINS, *Professor of Economics*, University of Miami
PETER E. ROSSI, *Assistant Professor of Econometrics and Statistics*, Graduate School of Business, University of Chicago
BRYNA SANGER, *Professor*, Center for New York City Affairs, New School for Social Research
EUGENE SMOLENSKY, *Professor*, Institute for Research on Poverty, University of Wisconsin at Madison
ROBERT M. SOLOW, *Professor of Economics*, Massachusetts Institute of Technology
ROBERT G. SPIEGELMAN, *Executive Director*, W.E. Upjohn Institute for Employment Research
WILLIAM J. SPRING, *Vice President, District Community Affairs*, Federal Reserve Bank of Boston
NANCY BRANDON TUMA, *Professor of Sociology*, Stanford University
DAVID WARSH, *Columnist*, The Boston Globe
WILLIAM L. WASCHER, III, *Economist*, Board of Governors of the Federal Reserve System
HAROLD W. WATTS, *Professor of Economics*, Columbia University
AARON WILDAVSKY, *Professor of Political Science*, University of California at Berkeley
ARNOLD ZELLNER, *Professor of Economics and Statistics*, Graduate School of Business, University of Chicago

THE FEDERAL RESERVE BANK OF BOSTON CONFERENCE SERIES

No. 1	Controlling Monetary Aggregates	June 1969
No. 2	The International Adjustment Mechanism	October 1969
No. 3	Financing State and Local Governments in the Seventies	June 1970
No. 4	Housing and Monetary Policy	October 1970
No. 5	Consumer Spending and Monetary Policy: The Linkages	June 1971
No. 6	Canadian-United States Financial Relationships	September 1971
No. 7	Financing Public Schools (out of print)	January 1972
No. 8	Policies for a More Competitive Financial System	June 1972
No. 9	Controlling Monetary Aggregates II: The Implementation	September 1972
No. 10	Issues in Federal Debt Management	June 1973
No. 11	Credit Allocation Techniques and Monetary Policy	September 1973
No. 12	International Aspects of Stabilization Policies	June 1974
No. 13	The Economics of a National Electronic Funds Transfer System	October 1974
No. 14	New Mortgage Designs for Stable Housing in an Inflationary Environment	January 1975
No. 15	New England and the Energy Crisis	October 1975
No. 16	Funding Pensions: Issues and Implications for Financial Markets	October 1976
No. 17	Minority Business Development	November 1976
No. 18	Key Issues in International Banking	October 1977
No. 19	After the Phillips Curve: Persistence of High Inflation and High Unemployment	June 1978
No. 20	Managed Exchange-Rate Flexibility: The Recent Experience	October 1978
No. 21	The Regulation of Financial Institutions	October 1979
No. 22	The Decline in Productivity Growth	June 1980
No. 23	Controlling Monetary Aggregates III	October 1980
No. 24	The Future of the Thrift Industry	October 1981
No. 25	Saving and Government Policy	October 1982
No. 26	The Political Economy of Monetary Policy: National and International Aspects	July 1983
No. 27	The Economics of Large Government Deficits	October 1983
No. 28	The International Monetary System: Forty Years After Bretton Woods	May 1984
No. 29	Economic Consequences of Tax Simplification	October 1985