

# *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 performance 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 trade-off 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 assem-

bled 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 Hecló 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; Hecló 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 Hecló, 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 ex-

perimental 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 pro-



posed 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 shar-

ing 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.