95th Congress
2d Session } COMMITTEE PRINT

Materials Related to Welfare Research and Experimentation

COMMITTEE ON FINANCE
UNITED STATES SENATE

RUSSELL B. LONG, *Chairman*

Assembled by the Staff of the
Committee on Finance for the Use of the

SUBCOMMITTEE ON PUBLIC ASSISTANCE

DANIEL PATRICK MOYNIHAN, *Chairman*

NOVEMBER 1978

Printed for the use of the Committee on Finance

U.S. GOVERNMENT PRINTING OFFICE
26-071 WASHINGTON : 1978

COMMITTEE ON FINANCE

RUSSELL B. LONG, Louisiana, *Chairman*

HERMAN E. TALMADGE, Georgia	CARL T. CURTIS, Nebraska
ABRAHAM RIBICOFF, Connecticut	CLIFFORD P. HANSEN, Wyoming
HARRY F. BYRD, Jr., Virginia	ROBERT DOLE, Kansas
GAYLORD NELSON, Wisconsin	BOB PACKWOOD, Oregon
MIKE GRAVEL, Alaska	WILLIAM V. ROTH, Jr., Delaware
LLOYD BENTSEN, Texas	PAUL LAXALT, Nevada
WILLIAM D. HATHAWAY, Maine	JOHN C. DANFORTH, Missouri
FLOYD K. HASKELL, Colorado	
SPARK M. MATSUNAGA, Hawaii	
DANIEL PATRICK MOYNIHAN, New York	

MICHAEL STERN, *Staff Director*
GEORGE W. PRITTS, Jr., *Minority Counsel*

SUBCOMMITTEE ON PUBLIC ASSISTANCE

DANIEL PATRICK MOYNIHAN, New York, *Chairman*

RUSSELL B. LONG, Louisiana	JOHN C. DANFORTH, Missouri
WILLIAM D. HATHAWAY, Maine	CARL T. CURTIS, Nebraska

CONTENTS

MATERIALS RELATED TO WELFARE RESEARCH AND EXPERIMENTATION

Over the past decade, the Congress has considered a variety of proposals for major changes in the Nation's welfare programs. During this period, a number of significant modifications of those programs have, in fact, been enacted into law while other proposed changes have failed to win the acceptance of the Congress. In either case, however, there has been a continuing high level of legislative interest in examining the welfare system and finding ways to improve it. Concurrent with (and, perhaps, because of) this legislative interest, there has occurred a very substantial amount of research activity related to welfare programs.

The most widely known of the research projects are a series of income maintenance experiments which provided various types of income guarantees to sample populations. The genesis of these experiments is described as follows in a three-volume study issued by the Institute for Research on Poverty, which was the prime contractor for the first of the experiments: "Briefly, OEO had advocated a national negative-income-tax program in 1965, had been unable to persuade the President to introduce the legislation, and, therefore, had decided to fund a project designed to produce hard evidence as to its feasibility." The first experiments were begun in metropolitan areas of New Jersey in late 1968 and continued for several years with the final report submitted in December 1973. Additional income maintenance experiments have been conducted in rural areas of Iowa and North Carolina, in Gary, Ind., and in Seattle and Denver.

These various experiments have provided data upon which numerous analytical studies have been based both by the individuals involved in the research projects and by other commentators. The conclusions drawn from the data produced by the studies have not always obtained universal acceptance. For example, the summary report on the New Jersey experiments published by the Department of Health, Education, and Welfare states: "The most striking feature of the findings is that the observed changes in labor supply in response to the experimental payments were generally quite small." By contrast, the data from the same experiment led John F. Cogan of the Rand Corp. to comment: "The central finding of an analysis of the labor supply response of white male heads of household in the New Jersey-Pennsylvania Negative Income Tax Experiment is a large, statistically significant labor supply withdrawal."

In addition to the highly publicized income maintenance experiments, there has been a considerable body of other research related to welfare programs over the past several years. These include evaluations of various aspects of existing programs, a longitudinal study of the welfare population, a number of experimental studies related to the employment of welfare recipients and studies of the administrative aspects of the programs.

The purpose of this document is to make available a selection of writings which summarize and provide some commentary on a number of the welfare research efforts which have been undertaken in recent years. Some of the studies reproduced have been copyrighted, and the committee has obtained appropriate permission to reproduce those items in this print.

I. The Dynamics of Welfare Dependency: A Survey

(By David W. Lyon)

A Summary of the Dynamics of Welfare Dependency: A Survey

The objective of this paper is to draw upon available research in welfare dynamics to answer a set of specific questions regarding the use of public assistance—questions that are central to the design of alternative income maintenance systems. We take stock of what we know and, by exclusion, what we don't know about the use of welfare.

First we look at length and pattern of dependency and answer the question: how is the caseload distributed between short-term and chronic users of welfare? Second, we describe the welfare decision and answer the questions: what are the major factors that bring a family to turn to public assistance? What are the reasons a family leaves the rolls? What impact do program features have on the welfare decision? Third, we investigate levels and sources of income for welfare families and answer the question: how much income from nonwelfare sources do welfare recipients have over time? We assess the impact of job training and employment programs on the welfare decision and attempt to answer the question: what effect would work programs have on improving chances for nonsubsidized employment and reducing the welfare rolls? Finally, we look at the long-term effect of dependency on family behavior and discuss the implications of research on caseload dynamics for welfare reform.

ORIGINS OF RESEARCH ON WELFARE DYNAMICS

Caseload forecasting for federal, state and local government welfare agencies.

Experimental studies of work incentives and income guarantees.

Studies of welfare families over time, identifying length of dependency, income sources and impact of government programs on the welfare decision.

This review emphasizes research on the AFDC and general assistance programs, drawing upon analysis of caseload behavior at the local, state and federal levels.

LENGTH OF DEPENDENCY

The existing welfare system provides cash and in-kind benefits that greatly exceed incomes available from full-time employment in minimum wage jobs.

Getting public assistance is not a permanent condition for a majority of families on the rolls: most cases stay on the rolls for less than 3 years and the average stay is between 2 and 3 years in duration.

Three types of caseload samples have been used to measure length of dependency: numbers on the rolls at a certain point in time; all cases ever on welfare; and first-time recipients (opening cohorts).

Each gives a different profile of the duration of stay and has a different purpose for policy analysts: over 60 percent of cases on the rolls at a point in time are cases of long-term continuous dependency (3 years or more); less than 10 percent over a 6-year span will be long-term cases; and a third of all cases in an opening cohort will be long-term dependents. Therefore, before conclusions are drawn from studies of welfare dynamics, the characteristics of the caseload sample must be known.

Patterns of welfare dependency suggest high levels of caseload turnover. Over half of the nearly 1,000,000 cases on welfare in New York City from 1967 to 1972 were replacing cases that had been on the rolls but had moved out of welfare.

THE WELFARE DECISION

Benefit levels vs. wages.—Findings support the alternative income hypothesis: as benefit levels or benefit/wage ratios rise, case openings, applications and welfare participation increase and employment rates decrease.

Benefit-loss rates.—Higher benefit-loss rates result in less work effort; lower rates result in more work effort. Lower benefit-loss rates tend to result in more mothers (AFDC) working rather than an increase in hours worked.

Policy variables like the benefit-loss rate are not likely to have much effect in moving families off the rolls. The mean employment rate—18 percent—is so low that large percentage changes in work effort do not change the fact that most AFDC mothers do not work.

Work incentives are likely to increase welfare costs because of higher administrative costs or higher benefits for mothers who already work. Lag in reporting income and caseworkers discretion in counting deductible income results in lowering benefit-loss rates.

Multiple benefits.—The AFDC grant is only 55 percent of the total income (welfare and nonwelfare) received by dependent families in New York City, Michigan and California. Eighty-three percent of all AFDC cases in New York City receive cash and in-kind benefits at a value higher than the poverty line, and 95 percent have multiple benefit incomes (food stamps, Medicaid, shelter allowances, child care, nonwelfare income and the basic AFDC grant) higher than that yielded by a full-time minimum wage job without government supplements.

Short-term cases had much higher levels of Medicaid-paid health care than long-term cases. Much of the movement onto welfare is caused by demand for health care not covered by private insurance plans. Welfare may mask a large number of families in need of low-cost health insurance rather than cash assistance.

Employment opportunities.—The job market has a measurable effect on the welfare decision in spite of widespread concern that public assistance is a system quite apart from the ups and downs of the national economy.

The explosion of the AFDC caseload during the 1960's, when the national economy was healthy, was related primarily to factors other than changes in employment opportunities.

Migration.—Rather than being a direct reason for interregional migration, the welfare system enhances the attractiveness of regions with high wage levels, because it represents insurance in case a job is not available. But the primary factor in deciding where to migrate

seems to be labor market conditions; differences in state AFDC benefit levels have only a minor influence on the relocation decision of poor families.

Family composition and desertion.—Reform of the welfare system is not likely to reduce the trend toward non-white female-headed families, a trend that has far more complex origins than the availability of public assistance.

Attitudes toward dependency.—Growth of the AFDC rolls during the 1960's was primarily the result of more people moving on the rolls and secondarily of increased grant levels. More people were enrolled because more people found out about its availability and because it was increasingly acceptable to be on welfare—both changes having been fostered by the welfare rights movement.

A similar change in attitude might greatly increase participation and caseload size for the AFDC-Unemployed Fathers program. Currently, in part because of sensitivities to income origin in this group, participation rates are low for this program.

Administrative factors.—Administrative discretion is an important factor in the constantly changing patterns of the AFDC caseload. The short-run forecasting of caseloads is confounded by sudden shifts in case openings and closings related to changing administrative procedures. However, it does not offset the evidence that welfare dependency is essentially an economic decision.

EMPLOYABILITY AND INCOME

Employability is an elusive concept because objective measures do not fully reflect either the potential for or motivation of welfare families for finding work.

Far more families on the welfare rolls are employable and receive earnings from employment over periods of 1 or 2 years or more than is reflected in point-in-time samples of the caseload.

The public cost of day care and related work expenses is so high for many welfare mothers that they may actually earn less than it takes to keep them employed.

Earnings play a major role in the long-term income package of families who are on welfare at various times, even though earnings are less important in those times when welfare is received.

There is little difference between the income package of female-headed families who have been on welfare and the package of those who have never been on welfare.

The income and behavior of welfare families must be tracked over time to understand how welfare fits into an overall package of family income.

EMPLOYMENT AND TRAINING PROGRAMS

There is little evidence that job training and employment programs have been the source of the dynamic patterns uncovered in studies of the welfare population.

Employment programs may, at best, decrease the level of public assistance payments while the duration of stay is unaffected.

CASE BEHAVIOR AND LENGTH OF DEPENDENCY

Welfare recipients may become used to dependency and more resistant to leaving the longer they are on the rolls. This phenomenon has been called the "settling-in" effect.

If settling-in occurs, it could result in the need for a continuation of stringent work and income eligibilty tests under any welfare reform option—from an incremental to a full guaranteed income approach.

There is no clear evidence on whether settling-in actually occurs and what, if any, policies are necessary to offset its effects.

IMPLICATIONS FOR WELFARE POLICY

It cannot be accurately predicted whether labor force withdrawal and/or work disincentives would be greater under a guaranteed income plan than under AFDC. The impact of a guaranteed income can only be partially estimated from existing research on welfare dynamics.

In spite of the many inefficiencies and inequities in the current income support program, most families use the welfare system as intended—as a temporary source of income during periods of unemployment or other loss of normal income.

INTRODUCTION

The picture of welfare dependency that has emerged since research on "the welfare decision" began in earnest over ten years ago is both clearer and more complex than the common image of a permanent welfare class. There is substantial evidence that families decide whether to apply for or remain on welfare in an economically rational fashion. As the dollar value of benefits increases, more families will choose welfare over their available income sources, and, as benefits are taxed away at higher rates there is less work effort. Still additional factors are at work. Analysis of welfare caseload dynamics (behavior of welfare families over time) has shown, for example, that the composition of a family strongly influences the length of a welfare stay. Intact households (AFDC-UF and general assistance cases) tend to stay on welfare for a shorter time than female-headed families (AFDC cases), and length of stay increases with the number of children in a case. Administrative changes have also been found to change movement off and on the rolls and, therefore, the average length of stay. The stringency of eligibility controls, employment reporting requirements and changes in recertification procedures have all had significant effects on the size of the welfare caseload. Finally, the attitude toward welfare of eligible families has affected their decision to seek assistance.

The objective of this paper is to draw upon available research in welfare dynamics to answer a set of specific questions about the use of public assistance—questions central to the design of alternative income maintenance systems. We take stock of what we know and, by exclusion, what we don't know about the use of welfare.

First, we look at length and pattern of dependency and answer the question: how is the caseload distributed among short-term and chronic users?

Second, we describe the welfare decision and answer the questions: what are the major factors that make a family turn to public assistance, what are the reasons a family leaves the rolls, and what impact do program features have on the welfare decision? Third, we investigate levels and sources of income for welfare families and answer the question: how much income from nonwelfare sources do welfare recipients have over time? We assess the impact of job training and employment programs on the welfare decision and attempt to answer the question:

what effect would work programs have on improving chances for non-subsidized employment and reducing the welfare rolls? Finally, we look at the long-term effects of dependency on family behavior and discuss the welfare reform implications of research on caseload dynamics.

Origins of Research on Welfare Dynamics

Interest in the dynamics of welfare dependency has stemmed from three quite different, but equally pragmatic, pressures to identify the effects of welfare policy and economic conditions on caseload growth. The first pressure was the desire by welfare administrators both at the federal, state and local levels to improve their forecasts of the caseload for budget purposes. A number of caseload models were developed to see if a careful sorting out of economic and policy factors would result in better predictions than resulted from "best guesses" by agency personnel. In general, the models do not give better predictions than simple trend analysis combined with good judgment. The models have been useful, however, in sorting out benefit levels, job opportunities and policy changes as each contributes to caseload growth.

A second area of research came as a response to proposals for simplifying the welfare system by means of a guaranteed annual income for all families. Congressional critics of a simplified system argued that productivity would be seriously and permanently dampened, with the result that many families might never feel the need to work again. Economists did not argue whether there might be some work disincentives—clearly there would be some—but they set out to see how much work effort declined when families received cash transfers of different amounts. The most notable study—the New Jersey Income Maintenance Experiment—was funded by the federal Office of Economic Opportunity and was designed to measure how work effort changed in response to different benefit levels and benefit-loss rates. For the first time low-income families were "tracked" over time to see how they responded to a regular flow of unearned income supplied by the government.

The third area of research on welfare dynamics stemmed from a lack on information on how families made use of welfare benefits over time. A substantial number of case openings and closings is recorded in any one month, but there is little information on how many of the cases closed were the same ones that opened in the same or subsequent months. In a broader context, policy analysts wanted to know how groups of low-income families use welfare and related transfer programs to supplement earnings, how frequently they use welfare over time, and what portion of annual income comes from welfare. The major policy question was whether the same families make recurring use of welfare without any other income sources, or whether welfare is more of a temporary, "backup" source of income when a family falls on hard times.

This survey draws upon research carried out in these three areas of investigation, concentrated primarily on the Aid to Families with Dependent Children (AFDC) program because it is the largest national program and has been the focus of most research efforts.[1] Reference is made to general assistance programs when appropriate.

---

[1] We have not incorporated detailed findings from the negative income tax experiments for two reasons: first, because they are widely known and, second, because family response to benefit levels and tax rates were generally consistent with the behavior found for families in the AFDC program.

Our main objective is to review what is known about the dynamics of the existing welfare system and what existing patterns of dependency tell us about how families might respond to modifications of the current system. We frequently draw upon research studies of welfare systems in New York City, California and other locations, in addition to studies of the national caseload. The reason is that very little information is available to researchers on the behavior of welfare families over time, and what is available tends to have been specially prepared for studies of city or state systems. A particularly rich set of findings is available for New York City and California—two of the largest local caseloads in the country. Whenever possible, we have compared the local findings with studies that used national data bases. On specific questions where comparisons are possible, local and national data give similar profiles. Where comparisons were not possible, the studies of local systems should be very suggestive of patterns at the national level.

There are a number of studies listed in the bibliography that are not referenced in the text. Many of these influenced our thinking and the research designs of the studies discussed, but it was simply not possible to draw them all into the presentation.

Length of Dependency

Concern that public assistance will create a permanently dependent population has been a criticism of both the existing welfare system and of proposals for reform. The concern is based on two arguments. First, that benefits are or will be so high that many families could have higher incomes from public assistance than from the paycheck typically accompanying low-wage jobs. Second, once on welfare there is little incentive to turn to work if $1 of welfare benefits is taken away every time earnings increase by $1. There is substantial evidence that the existing welfare system in some states does provide benefits that greatly exceed incomes available from full-time employment in minimum-wage jobs, and that at some earnings levels tax rates on multiple welfare benefits could be near or in excess of 100 percent. (See Lyon, et al., 1976; Smith, 1976; California-DBP, 1975; Aaron, 1973; and Hausman, 1975.) However, research on the dynamics of welfare dependency does not support the concern that receipt of public assistance is a permanent condition for a majority of families on the rolls. In contrast, researchers have found that most welfare cases stay on the rolls for less than 3 years; that the average stay is between 2 to 3 years in duration; and that there are far more cases on the rolls over a 5- or 6-year period than are reflected in monthly welfare statistics.

A common means for measuring length of dependency is to ask how long a sample of families drawn from the monthly caseload has been on the rolls since their cases were last opened. The national survey of the AFDC case load in January 1973 documents that 2 years is the median length of stay since the family interviewed last came onto the rolls. (See Table 1.)

There were more cases for 1 year or less—30 percent—than there had been for more than 5 years—18 percent. Two comparable samples of

AFDC welfare cases for New York City show a somewhat smaller proportion of short-term cases and more long-term cases than the national sample. Both New York City data files (from separate studies) show that the median duration of stay is about 3 years, or one year greater than the national median. None of the three studies suggests that most AFDC cases have a permanent attachment to the rolls.

TABLE 1.—AFDC CASES BY TIME ON WELFARE SINCE MOST RECENT OPENING, UNITED STATES (1973), AND NEW YORK CITY (1971 AND 1972)

Time on welfare since most recent opening	Percentage of all cases		
	United States, January 1973	New York City, Ostow/ Dutka, September 1971	New York City, Rand/ Rydell, December 1972
Less than 6 mo	17.3	6.0	20.4
6 mo to 1-yr	12.9	9.0	
1 to 2 yr	19.1	16.0	18.8
2 to 3 yr	15.4	12.0	13.7
3 to 4 yr	9.7	27.0	9.6
4 to 5 yr	7.2		10.0
Over 5 yr	18.4	30.0	27.5
Total	100.0	100.0	100.0
Median years since most recent opening	2.0	3.25	2.78
Number of cases in sample	31,000	873	249,000

Source: U.S. Department of Health, Education, and Welfare, *Findings of the 1973 AFDC STUDY: Part I*, National Center for Social Statistics, NCSS Report AFDC-1 (73), June 1974, p. 80; M. Ostow and A. B. Dutka, *Work and Welfare in New York City*, Johns Hopkins Press, Baltimore, 1975, table 3.1, p. 27; C. P. Rydell, et al., *Welfare Caseload Dynamics in New York City*, the Rand Corp., 1974, table 2.13.

THE EFFECT OF SAMPLE TIME FRAMES

The profiles in Table 1, however, can be misleading. The distributions of length-of-stay are for a point-in-time, month) sample and they do not reflect all previous periods of dependency. Welfare researchers have shown that point-in-time samples underrepresent short-term cases and that large portions of the caseload have intermittent rather than uninterrupted periods of dependency.[2] Significant evidence of the impact that sample time frames have on the analysis of welfare dynamics was presented by Rydell, et al. (1974). They compare the distribution

[2] For a discussion of the effect of time-frame censoring on the analysis of welfare dynamics, see J. R. Hosek's presentation in Appendix A of R-2002-HEW, "Multiple Welfare Benefits in New York City," by D. W. Lyon, et al., The Rand Corporation, 1976, I. 111.

of welfare histories over a 6-year period from 1967 to 1972 for cases on the rolls at a point in time (December, 1969) and for all separate cases on the rolls over the 6-year period. The analysis is for New York City welfare cases. Only 3 percent of the point-in-time sample were on the rolls less than 1 year, whereas nearly 50 percent of all cases on over the 6 years were on less than 1 year. (See Table 2.) In both caseload universes over 20 percent of the cases moved on and off the rolls intermittently over the 6-year study period, all other cases had continuous dependency from several months to 6 years in duration. So a longer time "window" on welfare dependency than one month suggests that there is very little true permanent dependency; only 6 percent of the "all cases on" universe were on the rolls continuously for 3 or more years compared to 63 percent for the same group in that point-in-time universe.

TABLE 2.—EFFECT OF DIFFERENT TIME FRAMES ON MEASURE OF WELFARE DEPENDENCY, NEW YORK CITY, 1967–72

Welfare history, 1967–72	Time frame of sample (percentage of all cases)		
	All cases on welfare (1967–72)	Cases on welfare at point-in-time (December 1969)	Cases in an opening cohort (February 1967 to June 1967)
Short-term continuous (less than 1 yr)	49.7	2.9	17.4
Intermediate-term continuous (1 to 3 yr)	23.4	8.9	10.8
Long-term continuous (over 3 yr)	5.7	63.2	33.3
Multiple-term	21.2	25.0	38.5
Short (less than 1 yr)		(.8)	(8.6)
Intermediate (1 to 3 yr)		(5.4)	(10.5)
Long (over 3 yr)		(18.8)	(19.4)
Total	100.0	100.0	100.0
Number of cases	390,000	391,000	65,900

Source: C. P. Rydell, et al., 1974, table 2.5 and 2.6.

The choice of a caseload universe clearly has an effect on conclusions drawn from research on welfare dependency. Point-in-time samples are the most frequently used simply because they are the easiest to draw from case records, yet long-term cases are larger and tend to receive

lower per capita welfare benefits than short-term cases. Research findings from point-in-time samples will give only one view of employment experience, income sources, length of stay and responsiveness to work incentives.[3]

Despite these shortcomings, point-in-time samples are useful in two ways. First, administrators and policy planners frequently want to know the characteristics of cases on the rolls "right now." In any one month most expenditures are made on cash payments to long-term cases, so a point-in-time profile of dependency accurately portrays how the budget is allocated. Second, very few data are available on dependency patterns for all cases on the rolls over long periods of time. No national statistics are maintained on case histories, and local social service agencies usually purge their machine-readable records of closed cases a month or two after payments are terminated. Two notable longitudinal data files of welfare cases—one for New York City (Rydell, et al., 1974) and another for Alameda County, California (Wiseman, 1976)—involved costly file sampling, searches and processing for even the most rudimentary case information. A number of states (among them Illinois, Michigan and California) are just beginning to keep complete case history records of all cases, including those that opened for only brief periods of time.

One survey of income dynamics, the Panel Study of Income Dynamics (PSID) carried out annually by the University of Michigan's Institute for Social Research provides the best available picture of dynamics for the national welfare caseload. Martin Rein used the PSID sample of 5,000 families to answer the question: "Is there a welfare class?" (See Rein, 1976.) He measured the duration of stay for every female-headed family that ever received welfare during the 7-year period from 1967 to 1973.[4] Like Rydell, et al., for New York City he found far fewer long-term cases than in a point-in-time sample.

Rein found 26 percent of those families ever receiving welfare to be long-term stayers with continuous dependency, and another 8 percent of long-term cases with intermittent dependency. (See Table 3.) In total, the chronically dependent cases were one-third of all families who had received welfare during the period. Nearly 40 percent of the sample had only short-term dependency of 1 year or less. Both the Rein and Rydell, et al., studies indicate that welfare is a "permanent" or chronic condition for a portion of the caseload, and that the portion is between one-third and one-tenth of all cases receiving assistance over a long time.[5]

---

[3] For a presentation of the differences between long-term and short-term cases, see Rydell, et al., 1974; and Lyon et al., 1976.

[4] Rein's female-headed families do not fully exhaust the AFDC caseload because he analyzes only women between 25 and 55 years of age. The 1973 AFDC Survey reports that half of the mothers on AFDC were less than age 30.

[5] The "all cases on" approach to measuring length of stay is complicated by an "end point" problem. Cases closing at the beginning of the sample time frame or opening at the end are counted as less than 1-year stays, even though they might be the end or beginning of a long stay. Rydell, et al. (1974), partially solves the problem by excluding cases in the beginning and ending months.

TABLE 3.—DURATION AND PATTERN OF WELFARE DEPENDENCY, THE UNITED STATES AND NEW YORK CITY, 1967-72/73

[In percent]

Welfare history	Rein-United States families ever on welfare with head 18 to 55 yr old, 1967-73	Rand/Rydell-New York City separate cases on welfare, all categories of assistance, 1967-72
Short-term continuous (1 yr or less)	38.0	49.7
Intermediate-term continuous (1 to 3 yr)	18.3	23.4
Long-term continuous (over 3 yr)	26.1	5.7
Multiple-term	17.7	21.2
Intermediate	(10.0)	.....
Long	(7.7)	.....
Total	100.0	100.0
Number of cases in sample	748	390,000

Note: Rein data is for families only with head between 18 and 55 years of age taken from the Panel Study of Income Dynamics; Rand data is for both family cases and single-person cases, including recipients of AFDC; AFDC-UF; general assistance (Home Relief); Aid to the Aged, Blind, and Disabled. Long-term dependency in Rein data is for 4 or more years; more than 3 in the Rand/Rydell data.

Sources: M. Rein, "Is there a Welfare Class," mimeo, 1976, table 2; C. P. Rydell, et al., *Welfare Caseload Dynamics in New York City*, The Rand Corporation, 1974, Table 2.5, p. 16.

A third approach to measuring the duration of welfare dependency is to track an "opening cohort" of cases over time. This segment of cases can provide answers to the question: "how long will welfare cases that open today stay on the rolls?" This approach is equivalent, for example, to assessing the length of stay of all patients admitted into a hospital on a given day—it controls for changes in the composition of the caseload over time. Again, Rydell, et al. (1974), shows the sharp difference between the welfare histories of cases in an opening cohort and all cases on the rolls at a point in time. A group of cases going on the rolls between February and June of 1967 and tracked over the subsequent 5.5 years had fewer long-term cases and more short-term cases than the point-in-time group. (See Table 2.) On the basis of the data in Table 2 for an opening cohort, a welfare administrator can say that of all cases getting on the rolls in any 5-month period, 42 percent will be on welfare for more than 3 years, either continuously (33 percent) or intermittently (19 percent).

Rydell, et al., also calculated how long it takes for one-half of all cases getting welfare for the first time on any given day a week to close—a concept they call the half-life of a welfare case. They found that AFDC families with two or more children have a half-life of 2.5 years, and those with only one child have a half-life of 1.5 years; with an average half-life of 2 years for all regular AFDC cases. In

short, having children seems to account at least in part for a family staying on the rolls longer.

A study of the AFDC caseload in the State of California by Boskin and Nold (1975) tracks the history of 440 cases opening in January 1965 over the subsequent 5 years—1965 to 1970. They estimate the mean duration of stay on the rolls to be strongly related to race and wage level expected by the household head in the job market. Whites were on the rolls an average of 1.5 to 2.0 years during the 5 years, and nonwhites averaged between 2.5 to 3.5 years over the period. Cases with heads expecting a higher than minimum wage in the job market have stays roughly 1 year less than cases with heads facing a lower than minimum wage, regardless of color. One significant point on this broad picture is that nonwhites facing lower than minimum wage job opportunities are expected to average nearly 5 years per stay on the rolls.[6]

CASELOAD TURNOVER

There is a great deal of turnover in the welfare caseload. Rydell, et al., found that there were 992,000 separate cases on the rolls in New York City between 1967 and 1972.

The highest monthly caseload in the 6-year period was 523,000 in September 1972, so there were nearly twice as many separate cases on welfare over the 6 years than there were in the month with the highest caseload. The authors conclude that "of the 992,000 cases on welfare in New York City from 1967 to 1972 . . . 226,000 were on the rolls at the start of the period, 297,000 came from caseload growth during the period, and 469,000 resulted from permanent caseload turnover."[7] Permanent caseload turnover occurs when a closed case never returns to the rolls and is replaced by a case never before on welfare. Thus, over half of the 992,000 cases on the rolls in the 6 years were simply replacing a case that had been on the rolls but left permanently. This profile gives further evidence of the high level of movement on and off the welfare rolls over time.

The Welfare Decision

Why this turnover rate? An obvious reason is that families move onto welfare when they have a sudden loss of income or incur such high expenses that they can't make ends meet. Critics of the welfare system have never been unwilling to accept the argument that welfare is, in principle, an economic decision; but a number of dramatic changes in the rolls during the last 15 years suggested that something other than a simple economic decision was controlling growth in the caseload.

First, there was the explosion of the rolls during the 1960's when the economy was strong. At a time when unemployment rates were low and the number of employed persons was reaching new highs, welfare dependency was expanding. Second, there frequently were wild swings in the number of monthly case openings and closings that could not be easily related to changes in the economy or changes in welfare benefit levels.

Welfare researchers spent a good part of the late 1960's and early 1970's sorting out the various factors that account for changes in the rolls. They found that, indeed, indigent families were quite rational

[6] See Boskin and Nold (1975), Table 3.
[7] C. P. Rydell, et al. (1974), p. 44.

in making a decision between work and welfare, but that their movement on and off the rolls was affected as much by personal attitudes toward welfare and changing administrative policies as it was by the level of benefits and the benefit-loss rate built into the programs. The basic decision to choose or not choose welfare was quite rational within the rules of the game, but as it turned out the rules were frequently changed.

While much work was done to show that the behavior of welfare recipients is consistent with the tenets of rational economic behavior, it also became clear that the same rationality might lead to periods of extended dependency simply because the AFDC program, in combination with other income transfer programs, actually tends to discourage taking a job (work disincentives). Until 1967 when the tax rate on earnings was lowered to 67 percent, AFDC recipients lost $1 of benefits for every dollar they earned—an effective tax rate of 100 percent. In the Congressional debate that focused on the pros and cons of the negative income tax, it came to light that AFDC families who also benefit from food stamps, Medicaid and housing subsidies could face tax rates as high as 120 percent on earnings under existing and proposed programs. The difficulty of lowering the combined tax rate on multiple program benefits was forcefully presented by Aaron (1974).

Other work disincentives were identified. Two of the more frequently cited were family dissolution and interregional migration: real or feigned abandonment of a family by a father could result in greater family income than if the household were to remain intact; and higher welfare payments in heavily urbanized States might encourage migration of poor families to central cities with high unemployment.

We turn now to an overview of research on welfare dynamics that has identified factors entering into the welfare decision and measured the contributions of each to the growth of the rolls.

ECONOMIC FACTORS

Benefit levels vs. wages.—A common test of the welfare decision as a rational economic choice is to compare potential benefit levels to wage levels from available jobs. After controlling for other factors, studies in this area generally support the alternative-income hypothesis: as benefit levels or the benefit/wage ratio rise, case openings, applications and welfare participation increase and employment rates among welfare families decrease. An excellent survey of the welfare models in the context of caseload forecasting is contained in Abrahamse, et al. (1976).

Two major conclusions can be drawn from efforts to trace the welfare decision as a choice between earnings from work or income from a welfare check. First, researchers have found the welfare decision difficult to trace (or model) and they have been appropriately cautious in stating their conclusions. The main reason is simply that there are not enough data to account for all the economic and administrative factors that enter into the welfare decision. Movement on and off the rolls is as much a function of the changing supply of welfare (eligibility rule changes, work test and administrative stringency) as it is of the demand for welfare.

Yet with all the complexities and inadequacies of research in this area taken into account, it is possible to say that benefit/wage relation-

ships go a long way in explaining welfare decisions by actual or would be recipients.

The second conclusion is that there is a definite tendency for models to become more elaborate when information on the welfare decision is incorporated into the analysis. The more elaborate models become, the less accessible they are to the staff of welfare agencies for forecasting purposes. Thus, there is a conflict between the desire of welfare administrators to know *why* the caseload is changing, and their need to know *what* will happen for use in annual budget estimates. Answers to the "why" question require researchers to develop more complicated models, while far simpler models may actually perform better figuring out the "what" of the future.

Sumrall, Bluestone, and others (1976) at the Social Welfare Regional Research Institute in Boston have concluded that quite different models may be required for different states and different regions of a state because of wide variations in recipient behavior and administrative policy. In effect, there does not seem to be a simple conceptual model of the welfare decision that can be uniformly applied to the welfare system. There are many systems (at least one for each state) and there may be more than one appropriate model to explain family behavior.

Benefit-loss rates.—A number of studies of the AFDC program have measured the effect of benefit-loss rates on work effort of case members. The most notable studies include Hausman (1970); Appel (1972); Garfinkel and Orr (1974); and Williams (1975). The benefit-loss rate is the rate at which benefits are taken away as earnings increase. Higher welfare rates are expected to result in less effort to get or stay on a job. The studies are fraught with the same data problems discussed above. But they also show on a broad basis, that welfare families do react in the expected way: higher rates, less work, lower rates, more work.

Appel (1972) analyzed the employment rate of AFDC mothers in Michigan before and after the institution of the lower tax rate on earnings. He concluded that the employment rate increased significantly throughout the state and in some cases almost doubled. However, he found no increase in earned income for the average case, suggesting that the lower benefit-loss rate encouraged more mothers to take some employment but that there was no indication that mothers on average worked more hours. The net result was that "short-run costs to the taxpayer increased for employed mothers who received child care subsidies because the actual earnings were so low." (Appel, 1972)

Garfinkel and Orr (1974), using a national data base for 1967, found that employment rates would increase by 7 percent for every 10 percent reduction in the welfare benefit-loss rate. Furthermore, they found that the employment behavior of AFDC mothers is quite similar to that of married women. However, they concluded that policy variables like the benefit-loss rate are not likely to have a big effect in families' getting off the rolls. Combining a series of major simultaneous policy changes, their estimates imply that 50 percent of the AFDC mothers in a typical state would still not work.[8] (Garfinkel and Orr,

[8] The changes would be to simultaneously decrease the guarantee by 40 percent, the tax rate by 35 percentage points and aggregate unemployment by 1.5 percentage points; increase the levels of set-aside and deductions by $50.00 and the percentage of the caseload receiving rehabilitation services by 20 percent, and impose a work test in all states.

1974, p. 283). Like Appel, they also concluded that work incentives may actually increase welfare costs because of higher administrative costs or higher benefits for mothers that already work.

Two studies of benefit-loss rates under the AFDC program suggest that rates may actually be below their nominal level. Barr and Hall (1975) found that effective tax rates were below 100 percent before the policy change in 1967. Lags in reporting earnings and flexibility in defining deductibles both lead to benefit-loss rates closer to 50 or 40 percent or even lower than the nominal level of 100 percent effective before the 1967 policy changes. With the drop to a nominal level of 67 percent in 1967, it is likely that the effective rate is now even lower. R. Williams (1975) estimated real benefit reduction rates for a different sample than Barr and Hall and found much the same thing—real tax rates were far below the nominal rate of 100 percent. In the 10 states sampled, the rates were consistently below 50 percent.

The Barr-Hall and Williams findings suggest that administrative discretion is extremely important to the maintenance of work incentives for welfare families. In fact, the institution of a guaranteed income plan with a 50 percent benefit reduction rate without work expense deductions would probably result in greater work disincentives than welfare families currently face under the AFDC program.

Multiple benefits.—On the other hand, work incentives reflected in low AFDC benefit-loss rates are offset by additional welfare benefits that are income conditioned—food stamps, Medicaid, public housing and day care. The combined receipt of benefits from these programs and AFDC have been shown to result in nominal benefit-loss rates in excess of 100 percent—a $1 increase in earnings results in more than a dollar loss in either in-kind or cash welfare benefit. (See Aaron, 1973.) W. Williams (1975) suggests that high marginal tax rates among various welfare programs may have been the major cause of the defeat of the Nixon Family Assistance Plan and, indeed, he notes that had the consequences of multiple program benefits been thought through, the administration might not have supported a negative income tax. There is no published measure of the actual marginal tax rate faced by AFDC families also receiving Medicaid, food stamps and day care, but clearly once a family loses eligibility for welfare, in-kind benefits may be curtailed in addition.[9] Thus, it is likely that work incentives built into the AFDC program are at least partially and may be fully offset by the existence of multiple program benefits.

The decision to turn to welfare and the duration of stay is likely to be affected not only by work disincentives, but also by the absolute level of benefits available from multiple program sources. A study by the Rand Corporation (Lyon, et al., 1976) of multiple benefit levels in New York City found that 83 percent of all AFDC cases received over $5,000 in cash and in-kind benefits during 1974; and that 95 percent of the cases were over an annual income ($4,200) that would be forthcoming from a job paying minimum wages.[10] A four-person case averaged over $6,600. The benefits measured included the AFDC grant, food stamps, Medicaid-paid health care, shelter allowances, child care and nonwelfare income. In total, the AFDC grant was found to be only 55 percent of the income available to the family.

[9] Shkuda (1976) found that many cases closed off the rolls fail to continue Medicaid-paid health care even though they are still eligible.

[10] A minimum wage job was used as a measure of what an AFDC mother might find in the job market. But in fact, some in-kind benefits are available to and used by many low-wage workers.

Clearly, the welfare decision is made in response to more than just the benefit levels and tax rates embedded in the AFDC program.

The Rand study found that short-term and multiple-time cases had much higher levels of Medicaid-paid health care than cases on the rolls continuously for 3 years or more. The authors conclude that a good deal of the movement onto the rolls is caused by a demand for health care not covered by private insurance plans, and that welfare may mask a large number of families in need of health insurance more than income maintenance. In any event, it is clear that welfare dependency is very much tied up with short-run demand for health care, and that there have been no effective attempts to identify where the two systems overlap and where they should operate independently.

Employment opportunities.—An alternative view of the economics of the welfare decision is simply that unemployment creates a demand for welfare. Wage levels and benefit-loss rates have only a secondary effect if a family head is without a job and welfare is an available source of steady income. Most researchers have viewed unemployment as just one of a number of economic factors that explain changes in the caseload. In effect, the number of unemployed is an approximation of the size of the pool of families eligible for welfare, and increases or decreases in unemployment result in changes in the caseload. Aggregate unemployment rates or levels, however, have been found generally to be insignificant variables in models of the welfare caseload. Yap (1973), Saks (1974) and Rydell, el al., (1974) are among the studies that found either weak or no relationships between unemployment levels and the AFDC caseload.

The direct link between unemployment and the welfare decision seems to be strongest for the general assistance caseload. Kasper (1968) concluded that ". . . differences in labor market conditions rather than differences in the level of average welfare payments seem to be the major explanation of interstate variation in the proportion of the states' population receiving general assistance" (page 109). Rydell (1974) and Lyon, Menchik and Blais (1976) modeled case openings and case closings separately and found that the New York City Home Relief and AFDC-Unemployed Father caseload significantly related to levels of unemployment, while they did not uncover the same significance for the regular AFDC caseload. Again, the significance of the unemployment variable is highly sensitive to the exclusion of other variables in the model. Intuitively, these findings made sense. Eligibility for the AFDC–UF program is dependent on recent employment experience, and general assistance supports unemployed families or individuals that are temporarily unable to find or hold a job. On the other hand, a mother with dependent children covered by the AFDC program may find herself without support from a spouse for reasons quite independent from changes in the economy.

Nevertheless, there was considerable dissatisfaction with the apparent lack of association between economic health of the nation and the AFDC caseload. The encouraging findings for the general assistance and AFDC–UF caseloads were offset by their negligible size and cost compared to the regular AFDC program. In response, an alternative version of the unemployment effect was advanced by Venti (1975) and Sumrall (1976). Their argument is that most unemployment data includes only those establishments covered by unemployment insurance, and welfare recipients tend not to work in those establishments. Second, they argue that "the unemployment rate may be too highly

aggregated to proxy labor market conditions specific to the potential welfare population." (Sumrall, 1976, p. 11.) Models of Massachusetts, Georgia, Upstate New York and Washington were designed to include benefit, wage, benefit-loss and administrative variables in addition to a set of variables reflecting employment in welfare relevant industries—those with high turnover employment or those requiring only low levels of skills training. They also include an agricultural employment variable and the unemployment rate for all industries. For every region modeled they find that one or more of the employment opportunity variables are significant, in addition to variables relating to benefit levels, tax rates and administrative factors. (See Sumrall, 1976, Table 1.)

At the same time the Sumrall/Venti work was underway, Harrison and Rein (1976) also began to uncover relationships between unemployment and welfare that were stronger than identified in earlier models. The Department of Benefit Payments in California was sufficiently interested in the relationship between the AFDC rolls and the economy to fund a study by the Rand Corporation solely for the purpose of identifying and describing the direct relationship between work and welfare, if any. In answer to the question "does the AFDC caseload depend directly on the California economy?" the Rand authors give a cautious yes, but conclude that the relationship is really quite complicated and suggest, by implication, that a rather formidable modeling effort would be required to relate the welfare decision to changes in the health of regional economics. (See Abrahamse, 1976.)

In conclusion, there is increasing evidence that the job market does have a measurable effect on the welfare decision in spite of widespread concern that "welfare" is somehow a system quite apart from the ups and downs of the national economy. By its very design, the AFDC program was not intended to be primarily an unemployment insurance program, but many of the factors that bring about eligibility for AFDC—desertion, loss of employment for women with children, and loss of nonearnings income sources—are likely to be on the rise in recissionary periods. The explosion of the AFDC caseload during the 1960's when the national economy was healthy was related primarily to factors other than unemployment or employment opportunities, but on the basis of existing research it is clear that the job market does affect the welfare decision and a healthy economy will be an essential ingredient in the nationwide reduction of welfare dependency.

Migration.—If welfare-eligible families respond to high benefit/wage ratios by getting on the rolls in large numbers, then many observers have suggested that states with high benefit levels probably attract poor families from low-benefit states. None of the studies on migration done since 1968 has found a strong link between state benefit levels and migration, but the analyses have not incorporated benefit/wage ratios. The best measures offered so far indicate that black migrants to a city are less likely to be poor or on welfare than black families born and raised in the largest cities (Long, 1974); and that new arrivals to New York City are less likely to go on welfare than those who have been in the City for several years (de Ferranti, 1974). Utilization of the welfare system appears to increase gradually the longer a family stays in New York City, suggesting a discouragement after arrival in the City rather than a motivation for moving to the City.

In a detailed study of Puerto Rican migration into New York, Maldonado (1976) concludes that high wage levels are even further enhanced by the welfare system "acting as an insurance against unavailability of jobs." Given that higher income states tend to have higher benefit levels (Orr, 1976) it is difficult to separate the direct wage effect from benefit levels. On the other hand, Reischauer (1971) concluded that patterns of black migration from the South, while primarily a function of labor market conditions, are significantly influenced by welfare opportunities in the cities of destination.

Holmer carried out a thorough review of research on migration and the welfare decision and reached two major conclusions. First, welfare-induced migration has not been an important direct cause of the expansion of the AFDC rolls. Second, "migration behavior is shaped primarily by regional variations in labor market conditions and differences in state AFDC policies have, at most, a minor influence on the residential location decisions of the poor." (Holmer, 1971, p. 33.) In effect, existing research suggests that job opportunities and wages—not welfare benefits levels—dominate the interregional migration decision made by poor families.

FAMILY COMPOSITION AND DESERTION

The welfare decision is affected by the size and composition of the eligible family. An extensive analysis of welfare participation rates and duration of dependency carried out for New York City throws considerable light on family structure as a factor influencing the welfare decision. (See deFerranti, 1974; and Rydell, 1974.) First, families eligible for welfare where both parents are present (AFDC–UF and general assistance eligibility) have considerably lower propensities to turn to welfare than female-headed cases (AFDC eligibility). Only 45 percent of eligible intact families were on welfare in New York City in 1970 compared to 83 percent of those families with an absent parent. These participation rates are comparable to recent figures for the national AFDC–UF and AFDC caseload.

Second, large families tend to have higher participation rates than small families. Only 30 percent of intact families with one or two children turned to welfare, while 60 percent of intact families with five or more children were on the rolls in 1970.[11] (deFerranti, 1974.) For female-headed families, the range in participation rates is narrow—from 78 percent for small families to 90 percent for large families.

Consistent with their higher propensity to go on welfare, AFDC cases in New York City had a longer median length of stay (2.0 years) than AFDC–Unemployed Father cases (0.7 years). (Rydell, 1974.) AFDC cases with two or more children had a median length of stay of 2.5 years compared to 1.5 years for one child cases. Thus, the size and composition of families influence the dynamics of the welfare decision in much the way expected. Female-headed families have a higher propensity to go on welfare, and they stay on longer, than intact families. Larger families have higher propensities for dependency, and they stay dependent longer, than smaller families.

[11] Part of the higher participation rate of larger families is caused by the AFDC benefit structure—more persons, more benefits. As was discussed above, higher benefit levels tend to result in higher participation rates.

The high number of female-headed families on welfare has frequently raised the question of whether the current welfare system actually encourages family dissolution. At first glance there does not appear to be any evidence for this claim: welfare-eligible intact families generally have lower participation rates than eligible female-headed families; and if welfare were desired, more intact families could simply choose to apply for AFDC–UF without desertion by the father. However, the same families could realize even more income when the father deserts with the mother applying for regular AFDC assistance, and the separate households sharing the combined income of welfare and whatever full-time or part-time work is available to the father. The availability of food stamps, Medicaid and child care to the children in the AFDC case makes desertion even more financially attractive. A number of researchers have attempted to test the hypothesis that economic incentives contained in the AFDC program lead to more family dissolution than would occur without the program. A study by Honig (1974) tends to support the hypothesis. Honig isolates the independent effects of welfare on family splitting and on AFDC recipiency rates among female-headed families. After controlling for a number of relevant socioeconomic variables, she finds that the size of the average AFDC benefit in 44 metropolitan areas has a significant effect on the proportion of adult women who are heads of families and on the welfare participation rate of these families. While there are some methodological problems involved, her findings are supported by two other studies.

Bernstein and Meezan (1974) interviewed 451 welfare mothers in New York City and asked them whether the availability of welfare influenced their decision regarding separation from a husband or friend—14 percent said it did influence their decision. Because of the vagueness of response to the question by some mothers, the authors suggest that the "true" proportion of desertions caused by welfare is more than 14 percent. Their conclusion is that at least 75 percent of the mothers had relationships marked by such a high degree of tension that separation was either justified or that the choice was simply not theirs to make. "While some of these intolerable relationships might not have broken up had it not been for the availability of welfare, it cannot be argued that their preservation would have been desirable for the participants, the children or society." (Bernstein and Meezan, 1975, p. 100.)

Ross and Sawhill (1975) in a seminal book on female-headed families conclude that welfare benefit levels have a modest influence on the proportion of nonwhite women who head families with children. However, they feel that a number of underlying behavioral responses to welfare, in addition to desertion, may be the cause. Other candidates offered include the bearing and keeping of illegitimate children; setting up of separate households by women who without welfare would have lived with relatives or friends; and the delay of remarriage. The authors feel that all of these behavior patterns are probably supported by the availability of welfare, but that reform of the welfare system is not likely to reduce the trend toward female-headed families, a trend that has far broader origins than the availability of public assistance.

ATTITUDES TOWARD DEPENDENCY

One of the more perplexing aspects of the welfare decision is the apparent change in attitudes towards welfare that occurred in the late

1960s. Boland's definitive article on participation in the national AFDC program concludes that participation rates among eligible female-headed families increased from 63 to 91 percent between 1967 and 1970. (See Boland, 1973.) She attributes most of the massive growth of the rolls during this period to the increased rate of participation, with increases in grant levels (resulting in a larger pool of eligibles) contributing a much smaller share to the growth. Why did participation suddenly become virtually 100 percent after years of apparently low interest in welfare benefits?

A persuasive answer is that increased knowledge about the welfare system combined with a legitimatizing of dependency, both fostered by the welfare rights movement, resulted in an outpouring of applications by female-headed families. The national welfare rights movement made an across-the-board attempt to inform low-income families of their right to benefits, and to simplify a cumbersome applications procedure. Jackson and Johnson (1974) estimated that in an 8-month period during 1968 in New York City, the AFDC participation rate rose from 53 to 62 percent. The increase was due largely to the direct effect of the welfare rights movement on applications and its indirect effect on lower closing rates and higher grant levels. Whether or not the shift in national participation rates was largely due to the welfare rights movement, most researchers agree that a dramatic change in response to the welfare system by eligible families occurred during the late 1960s.[12]

An answer to what really happened to cause the increase in participation of welfare eligible female-headed families during the late 1960s has become an issue with respect to the national AFDC-Unemployed Fathers program. Participation in the AFDC–UF program is estimated to be anywhere between 20 and 40 percent of all eligible families. (See Boland, 1973; Hollenbeck, 1975; Lidman, 1975; and Rein, M., 1972.) Some have argued that few families know they are eligible for the program; others that work tests and eligibility rules are sufficiently stringent to make the expected benefit level small and of short duration; and still others that welfare has a stigma for families with strong labor force contact, and that unless attitudes change participation rates will continue to be low.[13] Because AFDC–UF eligibility rules are so stringent, it could be that the eligible population has been overestimated. In any event, participation in the AFDC–UF program has been low since its inception in the early 1960s, and the rates remain low even during periods of economic recession.

The response of eligible families to program benefits has a direct effect on program budgets. Because little is understood about why families use or do not use the AFDC–UF program, proposed changes to integrate AFDC–UF and the unemployment insurance (UI) system will be subject to widely varying estimates of cost. If the UF program continues to be funded separately from UI in DHEW, and is administered by local welfare agencies, the budgetary consequences of relaxing the work test, for example might be quite small. However, if the program is operated through employment security offices of DOL, the attitudes of eligible families might change and a major increase in participation and cost could result just because state employ-

[12] Part of the response might have been related to the new availability of Medicaid-paid health care during the late 1960s. However, the major expansion of the Medicaid program did not occur until after 1970.

[13] The Rand Corp. is currently under contract to DHEW to assess which of these explanations is most appropriate and how policy changes in the AFDC–UF program would change participation.

ment service offices are associated with contributory benefit programs rather than welfare. Because attitudes of families eligible for AFDC–UF may be highly sensitive to the source of benefits and means for enforcement of eligibility rules, one new policy option may be considerably more costly than another, depending upon the organization of program administration.

ADMINISTRATIVE FACTORS

We turn now to an aspect of the welfare decision that falls outside the control of recipients—the setting and interpretation of administrative rules and regulations. Each state welfare department has an allowance schedule and benefit-loss rate that is set by legislation, but there are a series of deductibles, disregards, set-asides, work tests, certification and recertification procedures, employment referral requirements and employability definitions that offer a great deal of administrative discretion to local welfare agencies. Changes in these rules and regulations and the stringency with which they are enforced have a major impact on the welfare decision; the number of case openings, closings and transfers; and, consequently, caseload dynamics. Case dynamics are also influenced by administrative problems of keeping track of welfare families once they start receiving assistance. Families move within and between welfare jurisdictions frequently without notice of their changed address. Cases may be closed and reopened within 2 or 3 months simply because of problems in tracking the location of families.

One analysis that relates administrative actions to case dynamics is by Rydell (1974) and Quint and Brown (1973) for the New York City caseload in 1972. They found that administrative factors were a major cause of case closings but that many cases eventually reopened and continued to receive assistance because of administrative errors or misassessments. First, they looked at all AFDC case openings during 1972 and found that nearly half the reasons for openings were due to household changes (loss of father), 16 percent for unemployment or earnings loss reasons, and 17 percent for medical reasons—the type of factors we have covered above in discussing the welfare decision. (See Table 4.) However, another 20 percent of the openings were due to factors that suggest the case was on the rolls before and was being reopened without having lost its eligibility in the interim—these are "contact reestablished," "administrative" and "other" reasons. Cases that opened for the other reasons might also have been on welfare before, but their closings probably resulted from a clear loss of eligibility. In contrast, over 17 percent of case closings were for administrative reasons, with another 35 percent for reasons of "contact lost." (See Table 4.) A lot of the contact lost activity was likely to be families who moved within or out of the city without any forwarding address. Over 75 percent of closing activity resulted from direct action by welfare offices for reasons that had no clear relationship to income gains or change in household composition.[14]

[14] National data on case closing (discontinuance) reasons suffers from the same problem of aggregation into uninterpretable categories. A study by Ketron, Inc. (1973) of a national AFDC survey for 1969–1970 indicates that over 40 percent of all closed cases in a year are for "other" reasons that cannot be related directly to a specific eligibility change or administrative action.

TABLE 4.—PERCENTAGE DISTRIBUTION OF AVERAGE CASE OPENINGS AND CLOSINGS PER MONTH, BY REASON, AFDC CASES IN NEW YORK CITY, 1972

Opening reason	Percent of all openings	Closing reason	Percent of all closings
Unemployment	13.4	Employment	11.4
Reduction in nonwage income	3.2	Increase in nonwage income	3.0
Medical	17.5	Death	.2
Change in household composition	45.6	Change in household composition	8.5
Contact reestablished	7.7	Contact lost	35.3
Administrative	4.8	Administrative	17.5
Other	7.7	Other	24.0
Total all reasons	100.0	Total all reasons	100.0

Source: Rydell, et al. (1974), tables 5.6 and 5.7.

The researchers then measured the proportion of reopened cases in 1972 that were closed for what they called spurious reasons: closings that resulted from error rather than from change in client need. Operationally, a closing was called spurious if it ever reopened by reason of "reestablished contact," or if it reopened for administrative reasons within thirty days of closing. In both instances it was unlikely that a change in need ever occurred. During 1972, the researchers found that 56 percent of closed cases that reopened within 1 month were spurious closings, and 42 percent of all closed cases reopened within 11 months were spurious closings. (See Table 5.) Rydell, Quint and Brown concluded that a significant portion of short-run caseload dynamics is administratively induced—either through error or the difficulty of monitoring the frequent intra-city moves of low-income families.

TABLE 5.—REOPENINGS OF WELFARE CASES DUE TO SPURIOUS CLOSINGS, BY TIME SINCE CLOSING: NEW YORK CITY, 1972

Months between closing and reopening	Reopenings of cases closed in 1972	Percent of reopenings due to spurious closings
Less than 1 mo	8,451	68.1
1 mo	14,005	55.6
2 mo	5,455	21.7
3 mo	2,857	10.3
4 mo	1,867	7.8
5 to 11 mo	4,106	2.9
Total within 11 mo	36,741	41.6

Source: Rydell, 1974, p. 68.

An analysis of the eligibility control program in New York City by Lyon, et al. (1976) showed that a major decrease in the AFDC caseload from mid-1972 to early 1974 was attributable to a photo identication card and mail recertification program. Because of these two programs the caseload declined by 20,000 cases—a 6 percent decrease. Mailed requests to report to welfare centers for the photograph resulted in some confusion and, frequently, a failure to report. Persons failing to report without cause had their cases closed. Individuals not responding by mail to the recertification program were given warning notices and, if a response was still not forthcoming, the case was closed. The caseload declined simply because there was a higher monthly average of case closings than had ever been recorded. However, by mid-to-late 1974 case reopenings averaged nearly three times what they had been before the eligibility control programs were launched, and by early 1975 the caseload was back to the level it was in mid-1972. The authors conclude that many eligible cases were being closed in error as a result of the programs, and that they were reopening after a hiatus of 1 to 2 months.

Further evidence of the role of administrative discretion is provided by Wiseman (1976) in a study of the AFDC caseload in Alameda County, California. Wiseman drew a random sample of AFDC cases on the rolls over a 6-year period from 1967 to 1972, and developed detailed dependency histories for each of the cases sampled. He modeled the effects on job-taking of four different policy changes that had occurred in the treatment of earnings and other income when calculating grant levels. Rather than increasing the likelihood of job-taking as expected, the policy changes had no observable impact on the work behavior of the case members. He suggests that the County partially offset the effects of the new policies throughout the study period by "steadily tightening treatment of work expenses." (Wiseman, 1976, p. 67.) In other words, while broad policy was being implemented to increase work effort, the County was tending to discourage job hunting by allowing fewer and lower levels of work expenses.

In summary, a significant portion of the welfare decison as reflected in case openings and closings is affected by administrative factors. After reviewing the welfare caseload turnover literature Friedman and Hausman (1975) conclude that "if turnover is so much a function of program characteristics, economic and administrative, then any particular findings on turnover (dynamics) should be interpreted as having applicability limited to the specific program studied." (Friedman and Hausman, 1975, p. 24.) They even go so far as to question whether findings uncovered to date on the patterns of welfare dynamics can really be generalized to patterns that would occur if simpler grant schemes were implemented with less administrative discretion.

Our judgment is that administrative discretion does cause an important part of the dynamic behavior found in the AFDC caseload. The stringency with which rules and regulations are enforced can definitely affect the length of stay on the rolls in the short run (6 months or less), but over periods of 1 year or more, research evidence supports the view that the welfare decision is made in ways that are consistent with broad policy variables like benefit levels, and benefit-loss rates and alternative income sources as reflected in wage rates

and unemployment levels. The fact that the short-run forecasting of caseloads is confounded by sudden shifts in case openings and closings related to changing administrative procedures does not offset the evidence that welfare dependency is essentially an economic decision. The cost and caseload implications for new variations in the AFDC program can be estimated from the underlying behavior we find in the existing system(s). The wide variation in AFDC policy between states, in fact, provides an opportunity to estimate the impact of policies existing in some states on states without such policies. For example, the State of California has for years provided AFDC–UF coverage to families simultaneously receiving unemployment insurance coverage. Estimating the impact of the recent Corman bill on the national AFDC–UF caseload (permitting and requiring UI benefits be treated as nonwelfare income) will be greatly facilitated by California's long experience with an AFDC–UF program with just those features. This and similar policy options being considered for the AFDC–UF program (including mandatory state provision of UF benefits) can be evaluated directly from the experience of the existing AFDC program.

Furthermore, the evidence on length of welfare dependency and permanent caseload turnover discussed earlier suggests that welfare dynamics are related to a much broader range of factors that affect the income dynamics of a family over time. We turn to these income dynamics issues in the next section. But recall that Rydell (1974) found that nearly half of the 1 million cases on welfare in New York City between 1967 and 1972 represented permanent turnover of the caseload, i.e., they had replaced a case that had been on the rolls but left permanently. (See p. 11.) Also, Rein (1976) and Rydell (1974) found that between 40 and 50 percent of all families on welfare over 6-year periods were on the rolls continuously for only 1 year or less. While erractic administrative policies may generate a pattern of intermittent dependency for one case, it is evident that welfare is used quite differently depending upon the finances, household composition and health circumstances characterizing a family at a particular point in time. As a result, there is the wide but predictable range of dependency patterns shown in Table 3.

Employability and Income

A substantial effort has been made at the federal, state and local levels to determine the employability of welfare family members in order to encourage work effort. Employability has proved to be an elusive concept simply because objective measures do not reflect either the potential for or motivation of welfare families for finding work. The analysis of welfare dynamics has shown that far more families on the welfare rolls are employable and receive earnings from employment over periods of 1 or 2 years or more than is reflected in point-in-time samples of the caseload. Nevertheless, the public cost of day care and related work expenses is so high for many mothers on the AFDC rolls that they may actually earn less than it takes to keep them employed.

One approach to identifying employables on the rolls is to include adult case members already or recently employed, plus those adult case members who head a family with one or no children under the age of six. Studies of the New York City caseload by Lyon, et al. (R-1485) and the California caseload by Abrahamse, et al. (forthcoming) have used this approach.

In September, 1972, a full 50 percent of AFDC and general assistance recipients between 18 and 64 in New York City were found to be employable by this method, even though only 8 percent of the adult recipients were reporting earned income in the study month. Similarly for California, 60 percent of AFDC family heads in 1973 had no children under six or had current or recent work experience. This is a broad-brush approach to identifying employable welfare recipients, but it is generally consistent with the target population of the Work Incentive Program. To show how quickly the employability picture narrows if crude potential measures are incorporated, the Abrahamse study found that only 23 percent of AFDC heads in California had no children under six, had either recent work experience or some work experience, had a high school degree, and were in a prime age bracket for employment. In other words, the employable welfare population varies widely in numbers depending upon the purpose of program administrators. If they want to know how many welfare families might have a chance—no matter how small—of taking a job and reducing or eliminating their dependency, the broader definition can be used. If they want to target staff and employment support measures on families with higher probabilities of employment, the narrower definition is more appropriate. The Work Incentive Program is likely to be more effective in job placements, for example, if it targets on a narrower definition of employability.

An alternative approach is to identify the proportion of the caseload that would be capable of supporting itself independently of welfare, given wage levels that are commensurate with recipients education and occupation. Hausman (1969) found that 40 percent of AFDC adult recipients could support themselves independently of welfare if they had a job, but Warren and Berkowitz (1969) found this figure to be less than 20 percent for female family heads on welfare in California.

These findings were based on case behavior before the implementation of the monthly disregard of $30 and one-third of earned income. Lyon, et al. (R–1485) concluded that the wage rate paid to welfare recipients placed in jobs by the New York State Employment Service was not sufficiently high to eliminate dependency for most of those placed, although the placements did reduce the amount of the public assistance check. Findings about the relatively low attractiveness of wages compared to welfare payments are strengthened even more by the availability of multiple welfare benefits.

What, then, is the evidence about the attachment of welfare mothers and fathers to the labor force over time compared to their apparent employability at a point in time? Some of the best research to date on the income dynamics of welfare families has been done by Martin Rein using the University of Michigan's Panel Study of Income Dynamics (PSID). In a report entitled Sources of Family Income and the Determinants of Welfare (Rainwater and Rein, 1976) the authors identify female-headed families that were on welfare at least once over a 5-year period from 1967 to 1971. Some of the time over the 5 years there may have been a spouse present. First, the authors found that, in 1972, 41 percent of the female family heads were employed (they were not necessarily on welfare at the time of the 1972 interview), but that 69 percent were employed at least once during the 5-year period.

Second, they found that female-headed families have much larger portions of their income from earnings over the 5-year period than is

the case when they are on the rolls. Fifty percent of all income over the 5 years received by adult women heading households with children came from earnings, while only 24 percent in 1967 income came from earnings for those women on welfare in 1967. (See Table 6). The authors note that "female-headed families who received welfare at least once in the 5 years, but not during 1967, had an income package much like that of female-headed families that never received welfare." (See Rainwater and Rein, 1976). Their general conclusion is that earnings play a major role in the long-term income package of families ever on welfare, even though earnings are less important at the time welfare is received. (See Table 6.)

The authors carry their analysis of income/welfare dynamics further and examine the relative importance of family structure, the labor market and the welfare system to the likelihood of receiving welfare during the 5 years.

> The results clearly show that when we hold all other factors constant, family structure and labor market factors are about equally important in accounting for the years spent on welfare (from zero to five years). Within our labor market variables the likelihood of being employed and the number of hours worked, rather than wage levels, appear to play a more significant role in accounting for years on welfare. Those variables that *attempt* to measure welfare benefits and access to welfare do not appear to play as important a role in determining welfare status as do family structure and labor markets. (Italics mine.) (Rainwater and Rein, 1976, Chapter 7, p. 17.)

The Rainwater/Rein work again shows the importance of underlying economic features in the welfare decision, and the usefulness of tracking the income and behavior of welfare families over time to understand how welfare fits into an overall package of family income.

TABLE 6.—DISTRIBUTION OF INCOME BY SOURCE, 1967 TO 1971 AND 1967, FOR WOMEN AGED 24 TO 54 LIVING IN FEMALE-HEADED FAMILIES WITH CHILDREN [1]

[In percent]

Sources of income	Income from 1967 to 1971	Income in 1967	
		On welfare in 1967	Off welfare in 1967
Head's earnings	31.3	12.2	49.2
Spouse's earnings	3.9	.....	.....
Other earners	14.8	12.1	21.9
Assets	4.4	0.5	7.3
Private transfers	6.1	13.0	8.6
Welfare	29.0	55.7	.....
Work-related benefits	6.5	3.9	12.6
Food stamps	4.0	2.6	.4
Total	100.0	100.0	100.0

[1] The women had children in at least 1 of the 5 yr from 1967 to 1972 and were on welfare in at least one of the five years.

Note: Sample size is 523 for all women in female-headed families that were ever on welfare.

Source: Tables 15 and 19, Rainwater and Rein (1975) from the panel study of income dynamics, using sample of U.S. population.

Levy (1976) has done an extensive analysis of the same data file (PSID) to see how much movement there is by families in and out of poverty, rather than welfare. His findings are broadly consistent with those of Rainwater, Rein and Rydell for the welfare population. Levy identified a sample of people who were below the poverty line in 1967 and then tracked their income for the next 6 years to see how many stayed or moved above the line. He found that 47 percent remained poor through most of the subsequent 6 years; 26 percent moved out of poverty for 5 or 6 of the following 6 years; and 27 percent moved out of poverty for roughly half of the time between 1967 and 1793. Levy's analysis dispels the image of poverty as a permanent status for families, just as the work of Rein, Rydell, and Boskin and Nold does for welfare dependency.

Employment and Training Programs

How much of the dynamic behavior of the poverty and welfare population is attributable to the numerous employment and training programs funded by federal, state and local governments? The paper by Levitan in this series suggests that employment programs impact only modestly on the decision to leave welfare, and that job training programs benefit those who are most likely to find jobs anyway. Goldstein (1972) carried out an extensive review of job training programs for both welfare and poverty target populations and found some evidence that trainees had higher job placement rates and higher wages than control groups. But the findings were frequently inconsistent from study to study, positive effects were often small and methodological problems made firm conclusions impossible. Studies of the AFDC population that have findings generally consistent with these have been completed for the states of Florida, Michigan and Minnesota (Institute for Interdisciplinary Studies, 1972); New York City (Lyon, et al., forthcoming); and California (Wiseman, 1976), among others. In sum, there is little evidence that the effectiveness of job training and employment programs have been the source of the dynamic patterns uncovered in studies of the welfare population.

The dynamics of dependency are only affected when the chance of a case closing is increased by an employment program. A welfare recipient may be helped in finding a job, but still continue to be on the rolls with a somewhat lower assistance check. Wiseman distinguished between higher chances of a case termination and higher chances of taking a job when assessing the impact of employment services for AFDC cases in Alameda County, California. He found that employment services did not significantly increase the likelihood of welfare termination, but they did increase the likelihood of employment. (Wiseman, 1976, p. 68.) Thus, employment programs may, at best, decrease the level of public assistance payments while the duration of stay is apparently unaffected.

Rydell, et al. (1974) found that the proportion of long-term AFDC cases with employment income in New York City (7 percent) was just the same as for cases on the rolls 2 years or less. In fact, an even higher fraction of long-term general assistance cases had employment income (16 percent) than short-term cases (7 percent). They conclude that long-term dependency is as compatible with employment experience as is short-term dependency since many jobs simply do not provide sufficient income to eliminate dependency. "Partial or low-wage employment is not necessarily a stepping-stone to self-sufficiency, but is com-

patible with long-term dependency on welfare assistance." (Rydell, et al., 1974, p. 51.) After comparing a number of case characteristics (category of assistance, income sources, case size, employability, age of payee) with duration of stay and finding no one case type completely associated with one type of case history, the authors conclude:

> At the current state of knowledge about welfare dependency, chance events play a large part in determining whether a welfare case is a short-term or a long-term one. One can only talk of tendencies, or probabilities, that a case with given characteristics will have a certain type of welfare history. (Rydell, et al., 1974, p. 52.)

In addition, it is our assessment that employment and training programs have had a negligible effect on determining whether a case has short-term or long-term dependency.

Case Behavior and Length of Dependency

The single most important and recurring theme in discussions of welfare policy is the impact of cash payments on work incentives. Benefit levels, benefit-loss rates and employment programs have all been designed to encourage and cajole welfare recipients into employment—or, at least, not discourage employment by heavily taxing earnings. As we have seen, there is considerable evidence that, in general, welfare families respond to incentives as intended. In recent years, however, another aspect of the incentive issue has been raised that may prove to be equally important to the design of strategies for reducing work disincentives. It may be that families become more accustomed to a life of welfare dependency as each month passes, and are less responsive to work incentives after a year or more of dependency than they are when first turning to public assistance. Rydell has formalized this phenomenon into what he calls the "settling-in" effect—the behavior of welfare recipients may be conditioned by the length of time on the rolls. An earlier version of the settling-in phenomenon was an argument that welfare resulted in intergenerational dependency;[15] but, this version suggests that dependency on welfare strengthens within the life of one case.

There has not been an unambiguous test of the hypothesis that welfare cases settle in dependency, but both Rydell, et al. (1974), and Ketron, Inc. (1973 and 1974), present findings that are at least consistent with a settling-in phenomenon. They both calculate case closing rates for periods of 6 months, 1 year and 2 years after a cohort of cases has opened onto the rolls. If closing rates drop the longer a group of similar cases has been on the rolls, then there is some evidence that cases "settle" into dependency.

Rydell found that for all categories of assistance (AFDC; AFDC-UF; General Assistance; Aid to the Disabled, Aged and Blind [now SSI]) in New York City, closing rates for a group of cases opening onto the rolls at the same time decrease with case age. For example, AFDC cases with two or more children have a closing rate of 4.4 percent 3 months after opening; but for those cases that remain on the rolls from 3.5 to 5.0 years, the closing rate drops to 1.2 percent. In other words, the chance that an AFDC case will close right after opening is nearly four times higher than the chance after 3.5 years of dependency.

---

[15] See Greenleigh Associates, Inc., 1969.

Rydell found that using case age intervals of 3 months, 6 months, 1 year, 1.5 to 3.0 years, and 3.5 to 5.0 years, closing rates consistently drop with case age.

Similar findings have been presented in two separate reports by Ketron, Inc., using the national AFDC surveys for 1969, 1971 and 1973. They disaggregate case characteristics by age, sex, race and education of an AFDC payee (a considerably greater disaggregation than that possible in the New York City data used by Rydell) and conclude that cases "exhibit the tendency to remain on the rolls longer as the duration of case (age) increases." (Ketron, Inc., 1974, p. 5.)

Since the Rand/Rydell and Ketron, Inc. studies have been published, there has been considerable discussion and debate over the interpretation of decreasing closing rates with case age. (So far no written rebuttal of the two reports has been published.) [16] Critics argue that there are a number of alternative hypotheses that are consistent with case closing rates declining with case age. One hypothesis is that the two kinds of cases open onto the rolls—one type that will stay only a short time (movers) and another type that will stay a long time (stayers). Declining closing rates would also occur if this hypothesis were correct, because cases in the low closing rate group stay on welfare longer than cases in the high closing rate group. To date, there has been no additional research on case behavior that identifies and tests alternative hypotheses on how families adapt to welfare over time.[17] Until that research is done, Rydell felt comfortable with the conclusion that "there is currently no evidence to contradict the hypothesis that welfare cases become increasingly dependent on welfare the longer they stay on welfare." (Rydell, et al., 1974, p. 33.)

The notion that welfare cases may adjust to dependency and become less responsive to work incentives over time has important consequences for national income maintenance policy. Discussions of the New Jersey negative income tax experiment have emphasized that the brief duration of the project simply could not account for the long-term adaptive behavior of participant families. (See Aaron, 1975.) The findings that male participants did not make a significant withdrawal of work effort may not hold over a longer period of support. The policy options available to protect the economy from increasing program participation levels and gradual withdrawal from the work force have not been clearly identified for reform proposals under discussion to date. It may be that benefit levels and tax rates are inadequate policy tools by themselves. However, if stringent work and income eligibility tests are to be required, the administrative bureaucracy might be every bit as cumbersome, costly and inequitable as the current AFDC system. (See W. J. Cohen, et al., 1976.)

Implications for Welfare Policy

The evidence from research on welfare dynamics has mixed implications for policy reform. On the one hand, welfare recipients

[16] The first presentation of the Rand/Rydell settling-in hypothesis was in December, 1973, at a Welfare Research Conference in New York City sponsored by Administrator Jule Sugarman of the Human Resources Administration. It was subsequently presented at the December, 1975, Dallas meetings of the American Economic Association. (See Lyon, Rydell, and Menchik, 1975.)

[17] M. Wiseman and F. Levy at the University of California, Berkeley, do have some unpublished findings that number of months on welfare was insignificant in predicting closing rates for the Alameda County caseload. For a description of the model, see Wiseman, 70, 1976.

clearly respond in rational ways to benefit levels and benefit-loss rates built into federal and state AFDC and general assistance programs—higher benefits result in more participation; higher tax rates, less work effort. Further, the welfare system performs an important temporary income maintenance function. Forty to fifty percent of all families who turn to welfare stay on the rolls for 1 year or less; and less than a third of the average welfare family's income comes from public assistance when totaled over a period of 5 or more years. In other words, the majority of recipients use public assistance for income support in much the way other temporary support programs like unemployment insurance and workmen's compensation are used. Modifications to the means-tested programs are likely to produce responses by eligible families that are consistent with behavior under existing programs.

On the other hand, research findings from the existing welfare system have only limited transferability for describing behavior under a universal guaranteed income plan. Participation in a guaranteed income plan would be inclusive of nearly all low-income families, and the cost of participation in terms of administrative screening, income audits, periodic recertification and employability determination is likely, by design, to be considerably lower. It cannot be accurately predicted whether labor force withdrawal and/or work disincentives would be greater under a guaranteed income plan than under AFDC. The strong role of caseworker eligibility determination and management probably accounts for some of the dynamic patterns uncovered in the AFDC program, although researchers have uncovered little direct effect on overall dependency patterns from eligibility control and employment programs. The fact that each state welfare program lends itself to a somewhat different model of caseload behavior suggests that a unified national plan would have quite a different impact in different parts of the country—those impacts can be only partially estimated from existing research.

BIBLIOGRAPHY

Aaron, Henry J., *Why is Welfare so Hard to Reform?* The Brookings Institution, Washington, D.C., 1973.

———, "Cautionary Notes on the Experiment," in *Work Incentives and Income Guarantees*, J. A. Peckman and P. M. Timpane (eds.), The Brookings Institution, Washington, D.C., 1975.

Abrahamse, Allan, David deFerranti, Patricia Fleischhauer, and Albert Lipson, *Welfare Caseload Estimating Techniques: A Survey and Evaluation*, The Rand Corporation. R–1916–CDOBP, January 1976.

———. *AFDC Caseload and the Job Market in California: Selected Issues*, The Rand Corporation, forthcoming.

Ameniya, Takeshi and Michael Boskin, *Regression Analysis When the Dependent Variable is Truncated Lognormal, with an Application to the Determinants of the Duration of Welfare Dependency*, Stanford University, Institute for Mathematical Studies in the Social Sciences, Stanford, California, Technical Report No. 75, October 1972.

Appel, Gary Louis, *Effects of a Financial Incentive on AFDC Employment: Michigan's Experience between July 1969 and July 1970*, Institute for Interdisciplinary Studies, Minneapolis, Minnesota, March 1972.

Barr, Nicholas A. and Robert E. Hall, *The Probability of Dependence on Public Assistance*, MIT Working Paper No. 131, May 1974.

———, "The Taxation of Earnings under Public Assistance," *Economica*, November 1975, pp. 373–384.

Bernstein, Blanche, *Studies in Public Welfare—Income-tested Social Benefits in New York: Adequacy, Incentives, and Equity*, Paper No. 8, Subcommittee on Fiscal Policy of the Joint Economic Committee, Congress of the United States, 93rd Congress, 1st Session, Washington, D.C., July 8, 1975.

Bernstein, Blanche and William Meezan, *The Impact of Welfare on Family Stability*, Center for New York City Affairs, New School for Social Research, June 1975.

Bernstein, Blanche, Michael Rowan, and Anne N. Shkuda, *Obstacles to Employment of Employable Welfare Recipients*, Center for New York City Affairs, New School for Social Research, June 1974.

Boland, Barbara, *Studies in Public Welfare—Participation in the Aid to Families with Dependent Children Program (AFDC)*. Paper No. 12, Part 1, Joint Economic Committee, Congress of the United States, November 1973.

Boskin, Michael J. and Frederick C. Nold, "A Markov Model of Turnover in Aid to Families with Dependent Children," *The Journal of Human Resources*, Vol. X, No. 4, Fall 1975, pp. 467–481.

California State of, Bureau of Benefit Payments, *Total Value of Public Assistance Benefits in California*, 1975.

Cain, Glen G. and Harold W. Watts (eds.), *Income Maintenance and Labor Supply*, Institute for Research on Poverty Monograph Series, Academic Press, New York, 1973.

Chapin, Gene, "Unemployment Insurance, Job Search, and the Demand for Leisure," *Western Economic Journal*, Vol. 9, No. 1, March 1971.

Cohen, W. J., B. B. Conable, Jr., P. W. MacAvoy, and A. A. Ribicoff, *Welfare Reform: Why?* American Enterprise Institute, Washington, D.C., 1976.

de Ferranti, David M., Stephen Leeds, Joseph A. Grundfest, Valerie Leach, Peggy A. Parker, and Linda L. Prusoff, *The Welfare and Nonwelfare Poor in New York City*, The New York City-Rand Institute, R-1381-NYC, June 1974.

DeJong, Gordon F. and William L. Donnelly, "Public Welfare and Migration," *Social Science Quarterly*, Vol. 54, 1973.

Durbin, Elizabeth F., *Welfare Income and Employment: An Economic Analysis of Family Choice*, Praeger Publishers, New York, 1969.

———, "Work and Welfare: The Case of Aid to Families with Dependent Children," *Journal of Human Resources*, Vol. 8, Supplement, September 1973.

Feagin, Joe R., "America's Welfare Stereotypes," *Social Science Quarterly*, Vol. 52, March 1972.

Fechter, Alan and Stuart Greenfield, *Welfare and Illegitimacy: An Economic Model and Some Preliminary Results*, The Urban Institute, Washington, D.C., Working Paper No. 963–37, September 1973.

"Five Thousand American Families—Patterns of Economic Progress," a partial draft of Volume V, *Components of Change in Family Well-being and other Analyses of the First Eight Years of the Panel Study of Income Dynamics*, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, 1976.

Friedman, Barry L. and Leonard J. Hausman, *Work and Welfare Patterns in Low-income Families*, Brandeis University, Florence Heller Graduate School, Mimeo, June 1975.

———, *Work and Welfare Patterns in the New Jersey Experiment*, Brandeis University, Florence Heller Graduate School, Mimeo, December 1975.

Garfinkel, Irvin and Larry L. Orr, "Welfare Policy and the Employment Rate of AFDC Mothers," *National Tax Journal*, Vol. 27, No. 2, June 1974.

Greenleigh Associates, Inc., *A Pilot Study of Intergenerational Dependency in New York City*, New York, May 1969.

Greenston, Peter M. and C. Duncan MacRae, *A Diffusion Analysis of Participation in the Aid to Families with Dependent Children (AFDC) Program by States*, The Urban Institute, Working Paper No. 3606-07, Washington, D.C., July 1974.

Goldstein, J. H., *Studies in Public Welfare—The Effectiveness of Manpower Training Programs: A Review of Research on the Impact on the Poor*, Paper No. 3, Subcommittee on Fiscal Policy, Joint Economic Committee, U.S. Congress, November 1972.

Harrison, B. and M. Rein, *Some Microeconomic Relationships between Work and Welfare*, Department of Urban Studies and Planning, MIT, mimeo, 1976.

Hausman, L. J., "Potential for Financial Self-Support Among AFDC and AFDC-UF Recipients," *Southern Economic Journal*, July 1969, pp. 60–66.

Hausman, L., "The Impact of Welfare on the Work Effort of AFDC Mothers," *President's Commission on Income Maintenance Programs: Technical Studies*, 1970, pp. 83–100.

Hollenbeck, Keven M., *An Analysis of the Impact of Unemployment and Inflation on AFDC Costs and Caseload*, Mathematica Policy Research, Washington, D.C., February 13, 1975.

Holmer, M. R., *The Economic and Political Causes of the "Welfare Crisis,"* Ph. D. dissertation, MIT, June 1975.

———, *The Sensitivity of Transfer Payments to Changes in Economic Conditions and Public Policy*, HEW, ASPE, Office of Income Security Policy, Washton, D.C., partial draft, 1976.

Honig, Marjorie, The Impact of Welfare Payment Levels on Family Stability," in *Studies in Public Welfare (Part I), The Family, Poverty, and Welfare Programs: Factors Influencing Family Instability*, U.S. Congress, Joint Economic Comimttee, Paper No. 12, U.S. Government Printing Office, Washington, D.C., 1973.

Income Security for Americans: Recommendations of the Public Welfare Study, Report of the Subcommittee on Fiscal Policy of the Joint Economic Committee, Congress of the United States, 93rd Congress, 2nd Session, Washington, D.C., December 5, 1974.

Institute for Interdisciplinary Studies, *Final Report: AFDC Employment and Referral Guidelines*, Minneapolis, Minnesota, June 1972.

Jackson, L., and W. A. Johnson, *Protest by the Poor*, D. C. Heath & Company Massachusetts, 1974.

Kasper Hirschel, "Welfare Payments and Work Incentive: Some Determinants of the Rates of General Assistance Payments," *The Journal of Human Resources*, Vol. 3, No. 1, Winter 1968.

Ketron Incorporated, *Multivariate Regression Analysis of Annual National Turnover Rates Using the 1969 and 1971 AFDC National Surveys*, Wayne, Pennsylvania, August 1973.

Levine, Robert A. and David W. Lyon, "Studies in Public Welfare: A Review Article," *The Journal of Human Resources*, Vol. X, No. 4, Fall 1975, pp. 445–466.

Levitan, Sar, Martin Rein, and David Marwick, *Work and Welfare Go Together*, The Johns Hopkins University Press., Baltimore, Maryland, 1972.

Levinson, Perry, "How Employable are AFDC Women?" *Welfare in Review*, July-August 1970.

Levy, Frank, *How Big is the American Underclass?* Graduate School of Public Policy, University of California, Berkeley, 1976.

Lidman, R. M., *Why is the Rate of Participation in the Unemployed Fathers Segment of AFDC So Low?* Institute for Research on Poverty, Paper No. 288-75, Madison, November 1975.

Long, L. H., "Poverty Status and Receipt of Welfare among Migrants and Nonmigrants in Large Cities, "*American Sociological Review*, Vol. 39, February 1974, pp. 46–56.

Lyon, D., P. Armstrong, J. Hosek, and J. McCall, *Multiple Welfare Benefits in New York City*, The Rand Corporation, R-2002-HEW, August 1976.

Lyon, D., M. Menchik, and G. Blais, *Predictions of the New York City Welfare Caseload for 1975*, The Rand Corporation, R-1897-NYC, forthcoming.

Lyon, D., T. M. Palmerio, P. Armstrong, and R. Millar, *Employment and Job Training Programs for Welfare Recipients in New York City*, The Rand Corporation, R-1485-NYC, forthcoming.

Lyon, David W., C. Peter Rydell, and Mark Menchik, *Welfare Policy Research for New York City: Findings on the Dynamics of Dependency*, The Rand Corporation, P-5536, December 1975.

McCall, John J., *Income Mobility, Racial Discrimination, and Economic Growth*, D. C. Heath & Company, Lexington, Massachusetts, 1973.

Maldonado, R. M., "Why Puerto Ricans Migrated to the United States in 1947–1973", *Monthly Labor Review*, Vol. 99, No. 9, September 1976.

Meyers, S. M. and McIntyre, J., *Welfare Policy and Its Consequences for the Recipient Population: A Study of the AFDC Program*, Bureau of Social Science Research, Inc., December 1969.

Moynihan, D. P., "The Crisis in Welfare," *The Public Interest*, Vol. 10, 1968.

Ostow, Miriam and Anna B. Dutka, *Work and Welfare in New York City*, The Johns Hopkins Univesity Press, Baltimore, Maryland, 1975.

Orr, L. L., "Income Transfers as a Public Good: An Application to AFDC," *American Economic Review*, Vol. 66, No. 3, 1976.

Quint, J. and D. Brown, *Welfare Case Turnover in 1972*, City of New York, Human Resources Administration, Document No. 8857632-12, December 1973.

Rainwater, Lee and Martin Rein, *Sources of Family Income and the Determinants of Welfare*, Joint Center for Urban Studies of MIT and Harvard University, May 1976.

Rein, Martin, *Is there a Welfare Class?* Mimeo draft paper, 1976.

Rein, Mildred, *Conflicting Aims in AFDC-UF*, SWRRI #11, Social Welfare Regional Research Institute, Boston College, Massachusetts, August 1972.

Reischauer, R. D., *The Impact of the Welfare System on Black Migration and Marital Stability*, Ph.D. dissertation, Columbia University, 1971.

Rence, C. and M. Wiseman, *The California Welfare Reform Act and Participation in AFDC*, Working Paper No. 81, Department of Economics, University of California, Berkeley, 1976.

Ross, Heather L. and Isabel V. Sawhill, *Time of Transition: The Growth of Families Headed by Women*, The Urban Institute, Washington, D.C., 1975.

Rydell, C. Peter, Thelma Palmerio, Gerard Blais, and Dan Brown, *Welfare Caseload Dynamics in New York City*, The New York City-Rand Institute, R-1441-NYC, October 1974.

Saks, Daniel H., *Public Assistance for Mothers in an Urban Labor Market*, Industrial Relations Section, Princeton University, Princeton, New Jersey, 1975.

Schiller, Bradley, R., "Empirical Studies of Welfare Dependency: A Survey, *The Journal of Human Resources*, VIII-Supplement, 1973, pp. 19–32.

Sheehan, Susan, *A Welfare Mother*, Houghton-Mifflin Company, 1976.

Shkuda, Anne, *Former Welfare Families: Independence and Recurring Dependency*, Center for New York City Affairs, New School for Social Research, September 1976.

Smith, V. K., and G. A. Howitt, *The Economics Status of Michigan AFDC Families: An Analysis of Income and Benefit Receipt*, Studies in Public Welfare Policy, No. 8, State of Michigan, Department of Social Services, April 1976.

Studies in Public Welfare—Public Welfare and Work Incentives: Theory and Practice, Paper No. 14, A Staff Study prepared for the use of the Subcommittee on Fiscal Policy of the Joint Economic Committee, Congress of the United States, 93rd Congress, 2nd Session, Washington, D.C., April 15, 1974.

Sumrall, James, Jr., *An Overview of "Recent State AFDC Benefits and Caseload Dynamics*, "Social Welfare Regional Research Institute, Boston College, Chestnut Hill, Massachusetts, August 16, 1976.

Sweillam, Attia I., *Modeling the Public Welfare System: A Public Welfare Caseload Forecasting Model*, unpublished Ph. D. dissertation, Rensselaer Polytechnic Institute, Troy, New York, August 1974.

Venti, S. F., *The Determinants of AFDC in Massachusetts: An Econometric Analysis*, Social Welfare Regional Research Institute, Boston College, Massachusetts, Working Paper No. 25, May 1, 1975.

Walker, Neal and Luther Tweeten, "Determination of Participation Rates in Major Federally Subsidized Public Assistance Programs." *Rocky Mountain Social Science Journal*, Vol. 10, No. 3, October 1973.

Warren, Martin and Sheldon Berkowitz, "The Employability of AFDC Mothers and Fathers," *Welfare in Review*, July-August 1969.

Williams, Robert George, *Public Assistance and Work Effort: The Labor Supply of Low-income Female Heads of Household*, Industrial Relations Section, Princeton, New Jersey, 1975.

Williams, Walter, "The Continuing Struggle for a Negative Income Tax: A Review Article," *The Journal of Human Resources*, Vol. x, No. 4. Fall 1975.

Wiseman, Michael, *The Alameda County Welfare Sample: Graph, Tables and Stories*, Working Paper No. 80, Department of Economics, University of California, Berkeley, 1976.

———, *Change and Turnover in a Welfare Population*, Working Paper No. 70, Income Dynamics Project, Department of Economics, University of California, Berkeley. 1976.

Yap, Lorene Y. L., *A Gross Flows Model of AFDC Participation*, The Urban Institute, Washington, D.C., Working Paper No. 963–31, September 1973.

II. Excerpts From.—Work Incentives and Income Guarantees: New Jersey Negative Income Tax Experiment*

Introduction and Summary

(By Joseph A. Pechman, and P. Michael Timpane)

The New Jersey experiment in income maintenance was unique in several respects. First, it was the forerunner of numerous other large-scale, controlled social experiments testing the effects and feasibility of new social programs by observing how they operate in practice. Second, it was concerned with a live policy issue that could not be resolved satisfactorily on the basis of available data: whether cash allowances, whose net benefits decline as work income increases, significantly reduce work by the recipients. Third, the data to be collected would provide an unusually rich source of information for analysis by economists and other social scientists. This information was the basis for an exhaustive report that was submitted to the U.S. Department of Health, Education, and Welfare in late 1973 and early 1974.[1]

The experiment was conducted at a time when the public and policymakers alike increasingly were concerned about the cost of growing welfare rolls and the alleged impact of the welfare system on the incentives of the poor to lift themselves out of poverty. The traditional welfare system was designed to assist those who were not able to support themselves because of age, disability, or special family circumstances. Families headed by able-bodied men generally are excluded from such programs on the presumption that the head should support his family by working. In addition, the income of welfare recipients was taxed, often substantially, so that they were able to keep little if any of their income. Before 1967 the nominal tax rate of the federally financed assistance program was 100 percent of such income; thereafter, federal law allowed recipients to keep the first $30 per month plus one-third of any additional earnings, but supplementary benefit programs enacted by some state legislatures produced considerable variation in actual tax rates.

The negative income tax idea was advanced, first, to improve the traditional welfare system by providing a minimum, nontaxable allowance to all families; and, second, to maintain the work incentives of the poor who were able to work by permitting them to keep a significant fraction of their earnings. For example, the basic allowance might be $1,500 for each adult and $500 for each child, so that a family of four would receive a total of $4,000. Another possibility would be $1,000 for each person in the family regardless of age, which also would provide a total of $4,000 for a four-person family. As well as receiving

---

*Published 1975 by the Brookings Institution.

[1] Harold W. Watts and Albert Rees (eds.), *Final Report of the New Jersey Graduated Work Incentive Experiment*, vols. 1, 2, 3 and David N. Kershaw an Jerilyn Far (eds.), vol. 4 (University of Wisconsin—Madison, Institute for Research on Poverty, and Mathematica, 1973, 1974) (referred to later as *Final Report of the New Jersey Experiment*).

For a summary of the major findings by several of the authors of the report, see *Journal of Human Resources*, vol. 9 (Spring 1974), pp. 156–278. The term "New Jersey experiment" used throughout this book, in keeping with the title of the final report, includes the Pennsylvania experiment, as do the data in the report.

the basic allowance, the family would pay tax on each dollar of its income.[2] The rate of this tax might be 30 percent, 50 percent, 75 percent, or even 100 percent. Thus, with a basic allowance of $4,000 and a tax rate of 50 percent, a family with an income of $2,000 would have a disposable income of $5,000: $4,000 of the basic allowance and $1,000 of its own income left after the payment of taxes. With an income of $4,000, the family would have a disposable income of $6,000: the $4,000 basic allowance plus $2,000 of its own income after taxes. With an income of $8,000, a family would break even—the $4,000 basic allowance would exactly equal the tax of $4,000—and its disposable income would be $8,000.

The idea of a negative income tax met with considerable resistance, partly because of costs but primarily because many people believed that the guarantee of a minimum level of living would provide an irresistible inducement for a significant number of persons, especially able-bodied males, to reduce their hours of work or to stop working entirely. Although economists had been making econometric estimates of the effect of a negative income tax on work effort, the results had proved inconclusive.[3]

Aside from labor response, several other questions about the operation of a system of cash allowances could be answered by such an experiment. Of primary concern were the administrative problems of paying cash allowances. How should income be defined? Over what period should it be measured? Are there unusual problems of enforcement and compliance? Although these questions seem to be caught up in minor details, they are crucial to an evaluation of the workability and cost of a negative income tax.

Another set of questions concerned the effect of cash allowances on the life style of the recipient families. Will the families use the cash allowances for such frivolities as gambling, drinking, excessive entertainment, and other nonessential or harmful forms of consumption, or will they use the allowances to buy more or better qualities of such essentials as food, clothing, and shelter? What effect, if any, would the allowances have on the education of the children, the physical and mental health of the family members, and leisure activity?

The conference sponsored by the Brookings Panel on Social Experimentation on April 29 and 30, 1974, focused on these and related issues of importance to national policy. The remainder of this chapter summarizes the formal papers presented at the conference, as well as portions of the discussion that followed.

PLANNING THE EXPERIMENT

The opening paper, by Robert A. Levine, who was assistant director for research, plans, programs, and evaluation of the Office of Economic Opportunity (OEO) when the experiment was funded, traces the development within the government of the idea of commissioning a negative income tax experiment. To some, the major objective was

---

[2] Deductions might be allowed for such items as medical expenses, work clothes, union dues, and the like, but no personal exemptions would be allowed as in the positive income tax system.

[3] See Glen G. Cain and Harold W. Watts (eds.), *Income Maintenance and Labor Supply: Econometric Studies* (Markham, 1973). The econometric studies generally found income and substitution effects of the predicted positive sign, but the magnitude of the effects varied considerably.

experimental—to observe the labor-supply response of adult males, which they hoped would be slight—whereas to others it was mostly to demonstrate the administrative feasibility of a negative income tax. Levine's paper makes clear that the experiment was conceived by the OEO as part of a broad strategy to obtain administrative and congressional approval of a negative income tax.

The next paper, by Felicity Skidmore, describes the general design features of the experiment and the important decisions that were made in designing and carrying it out after the OEO agreed to fund the experiment and approved it in general terms. Skidmore's paper gives a detailed account of the problems that were encountered and of the administrative and analytical skill that was needed to overcome them. It was generally agreed at the conference that the experiment was carried out efficiently and that the analysis was done objectively and competently.

Skidmore's discussion reveals shortcomings as well as strengths in the experiment's design. The experiment benefited from important advances in techniques of field experimentation, notably the Conlisk-Watts design,[4] which enabled the experimenters to obtain for the same cost a significantly larger sample of relevant data than would have been possible with a simple random design. Despite problems of sample attrition and occasional disturbing influences of outside events, the basic design enabled the experimenters to collect a body of high-quality longitudinal data on the economic and social behavior of the working poor. At the same time, it was necessary to make some hard choices that limit the analytical uses of the data. These included choices with respect to the selection of the sample, the guarantees and tax rates, the site of the experiment, and the duration of the experiment.

The experiment concentrated on poor and near poor—that is, 125 percent or less of poverty income—male-headed families. This decision was based on a careful assessment of the policy significance of the behavior of this group under a negative income tax, but it excluded from study another policy-significant group, poor families headed by females; it led to an underrepresentation of families with full-time working wives; it excluded working families slightly above the eligibility line whose labor force participation also would be discouraged by a negative income tax; and it concentrated on intact families who differed in both known (for example, more children) and unknown (for example, emotional stability) ways from fractured families.

The experiment examined eight combinations of guarantees and tax rates. This design afforded an opportunity to observe the distinct effects of these two variables, but it also increased the required sample size and costs and thus limited alternatives concerning, for example, duration and number of sites. The samples for each treatment group, however, were all quite small.

The experiment operated at three urban sites in New Jersey and one in Pennsylvania. The sample is a reasonable representation of the eligible population, with balanced social and ethnic characteristics, in

---

[4] The full model is described in John Conlisk and Harold Watts, "A Model for Optimizing Experimental Designs for Estimating Response Surfaces," in American Statistical Association, *Proceedings of the Social Statistics Section, 1969*, pp. 150–56. Its application is discussed further in Charles E. Metcalf, "Sample Design and the Use of Experimental Data," *Final Report of the New Jersey Experiment*, vol. 2, pt. C, chap. 5. The design basically called for eligible families to be stratified by previous income; a predetermined number from each stratum was then assigned to the various experimental plans and the control group, with individual assignments achieved randomly.

the urban Northeast, but it is not representative of the entire country. Clearly, any limited site experiment reflects only imperfectly the conditions that a universal program would introduce.

The experiment was conducted for only three years in each site. A major drawback of the experiment is that the effects of a permanent negative income tax program remain unknown.

LABOR SUPPLY RESPONSE

The paper by Albert Rees and Harold W. Watts, who were codirectors of the experiment, summarizes the labor supply findings of the experiment. The major finding is that there was only a small (5 or 6 percent) reduction in average hours worked by the male heads of the families who received negative income tax payments. For reasons that are still not understood, this occurred entirely among white men; for black men, the response was insignificant but (surprisingly) positive, whereas for Spanish-speaking men, the response was also insignificant but negative. For white and Spanish-speaking working wives, who had a low participation rate to begin with, the negative reaction was greater—about one-third of previous work effort for whites and more than one-half for those who are Spanish speaking. The behavior of black working wives was not affected.

These results are evaluated in two papers by economists who had no connection with the experiment. Henry J. Aaron focuses much of his attention on the difficulties of interpretation arising from the fact that the welfare system of New Jersey was altered while the experiment was underway. He suggests that this change in the welfare system created complex and shifting incentives for individual families that are virtually impossible to quantify for purposes of analysis. Aaron also suggests other measurement problems that might lead to an underestimate of labor supply response. On the other hand, Robert E. Hall, who develops a model to evaluate the data on labor supply response obtained from the experiment, argues that the results are consistent with theoretical expectations and are statistically significant.

Nevertheless, a number of the conferees were not persuaded that the experimental data can be regarded as definitive on this score. Three major problems accounted for most of this skepticism. First, the effects noted by Aaron and others of changes in the New Jersey welfare system after the experiment was underway altered the net guarantees and tax rates in a haphazard way. Second, the limited time period covered by the experiment makes it difficult to draw conclusions about the effect of a permanent negative income tax. Third, the results for the nonwhite families included in the experiment, which are puzzling, may be unreliable because of high attrition rates.

Effect of the Changes in the New Jersey Welfare System

At the outset, it was considered essential to conduct the experiment in a st te in which male-headed families were not eligible for any welfare assistance. Only in such a situation would the difference between the families in the experimental and control groups reflect the effect of the negative income tax alone. New Jersey was chosen as the major site partly because it had no plan under the unemployed parent part of the federally supported Aid to Families with Dependent Children (AFDC–UP) program, which extended aid to unemployed fathers. (Male heads were eligible for the locally financed general assistance program, but the benefits under this program were extremely low.)

On January 1, 1969, however, New Jersey introduced a generous AFDC–UP plan for which most of the families in the experimental sample were eligible. Since families were kept in the sample, but they were required every payment period to choose between payments from welfare or the experiment.[5]

The change in the New Jersey welfare system had two effects on the results of the experiment. First, most of the families who originally were placed in two of the least generous experimental plans—those with a minimum guarantee of 50 percent of the poverty line income and a 50 percent tax rate (the 50–50 plan) and a 75 percent guarantee and a 70 percent tax rate (the 75–70 plan)—either chose to receive welfare payments or had incomes that exceeded the breakeven levels (see the explanation above). The few families that chose the experimental payments under these plans were dropped from the analysis of the labor supply responses because their number was so small. In addition, relatively few of the families assigned to the somewhat more generous 100–70 plan had incomes below the breakeven level and thus most received no payments. The experiment therefore provides little basis for judging the amount of labor that might be supplied by families under a 70 percent negative income tax plan. Moreover, the information from the remaining plans is not sufficient to measure the effects of different guarantees and tax rates.

Second, the availability of the welfare option meant that families in the experiment who were eligible for welfare faced very different guarantees and tax rates under each negative income tax plan from those faced by families ineligible for welfare. The result is that the observed differences between the experimental and control families cannot be related unambiguously to the stated guarantees and tax rates even for tax rates below 70 percent. As Aaron explains in his paper, the net guarantees and tax rates after taking into account the welfare option are very much lower—but not in the same proportion—for each of the plans than the stated guarantees and tax rates suggest.

The implications of this point were discussed at length by the conferees. One group pointed out that, because of the existence of other government assistance programs (in kind as well as in cash), it would have been impossible to arrange an experiment in which the net guarantees and tax rates for each experimental family were known with certainty. Under the circumstances, the experimental results should not be interpreted as providing estimates of the labor-supply response to any particular guarantee or tax rate, but rather of the response to the *difference* between the guarantees or tax rates among various plans. Such estimates still could be regarded as significant because any negative income tax plan probably will be superimposed on other existing programs, creating a complex environment that inevitably will alter the marginal rate that would actually apply to particular families.

---

[5] The Aid to Dependent Children (ADC) program was established by the Social Security Act of 1935 to provide income-related transfer payments to families whose father had deserted or divorced the mother. Therefore, it was a program restricted to families with female heads. The name of the program was changed later—with no implications for the functioning of the program itself—to Aid to Families with Dependent Children (AFDC). An amendment to the Social Security Act in 1961 expanded the program to cover certain categories of two-parent families by adding an "unemployed parent" segment to the AFDC program, thereafter called AFDC–UP. Because the unemployed parent invariably is a father, this program is also referred to as AFDC–UF.

The AFDC–UP (or AFDC–UF) program is at state option regarding decisions as to whether it will sponsor a program at all and as to the level of benefits and the eligibility criteria. New Jersey chose not to have such a program until January 1, 1969, when it instituted one of the most generous in the country. The state treasury could not support the program at that level, however, and its generosity was cut substantially in July 1971.

Another group felt that, for the reasons mentioned by Aaron, it is virtually impossible to infer anything about the marginal effects of the different guarantee levels or tax rates on labor supply. Because the net guarantees and tax rates are lower than the stated guarantees and tax rates, the calculations from the experimental data may understate the labor supply response to the type of negative income tax plans that are offered as policy alternatives in the United States. The combination of uncertainties as to the effect of different plans also makes it impossible to estimate the costs of various national negative income tax plans on the basis of the experimental results.

Short-Term Versus Long-Term Effects

Although the objective of the experiment was to estimate the labor supply response of a permanent negative income tax, to simulate the effect of a permanent plan obviously was impossible in an experiment lasting only three years. Early in the experiment, the study designers urged the OEO to cover some families for a period of three years and others for a period of five years, which would have permitted them to build a duration variable into the statistical analysis. But because of budget constraints, the OEO did not provide the necessary funds for a five-year period.

The temporary nature of the experiment is important because the work behavior of individuals may depend not only on their current needs and economic circumstances, but also on their future expectations. Whether the labor supply response in the experiment was greater or smaller than the response would be in a permanent plan is not clear. Some workers—particularly women who work to supplement their husband's income—might be induced by the negative income tax payment to stop working temporarily or to reduce their hours of work with the expectation that they could go back to work or work longer hours after the negative income tax arrangement ceased. This factor would lead to exaggeration of the labor supply response in an experiment of short duration. On the other hand, the enactment of a permanent negative income tax might bring on effects—for example, early retirement or less inhibition in quitting jobs—that a temporary experiment would be unable to capture.

Michael Boskin and Jacob Mincer, two of the formal discussants, as well as a number of other conferees, expressed the opinion that the temporary nature of the experiment probably led to an understatement of the labor supply response by male heads of families to a permanent negative income tax. Such workers usually are attached to a job that requires them to work a fixed number of hours a week. It is unrealistic to expect many of these workers to seek another job with shorter hours in the interest of obtaining a cash assistance payment for a three-year period, although over a longer period of time, employment structures might well change to make such decisions feasible. Moreover, the fact that the experimental families purchased relatively more durable goods than control families suggests that the former did not react to the additional income provided by the experiment in the same way that they would respond to a permanent change in income. Such behavior is consistent with the hypothesis that workers consider longer time horizons than three years in making economic decisions.

Although admitting that the experiment could not reproduce the conditions that would prevail under a permanent negative income

tax, other conferees took the position that the results are not necessarily biased in one direction or the other. Furthermore, it is not at all certain that the three-year duration was too brief to capture most of the labor supply effects of a permanent negative income tax. Whether the time period that controls economic behavior is three years, ten years, or a lifetime is still not known. In any event, the results of the experiment were considered to be important to the policymaker, at least as an indication of the short-run consequences of a negative income tax devision. Although the long-run effects remain uncertain, the experiment suggests that there would be sufficient time to readjust the terms of any plan that is actually adopted before the full consequences of the long-run effects become evident.

The Racial and Ethnic Puzzle

A major problem mentioned by both the experimental analysts and their critics was the unexpected behavior of the black (and, to some extent, the Spanish-speaking) workers in the experiment. On the average, little change occurred in the work effort of these participants in the experiment; in fact, as compared with the control groups, the work effort of black males actually increased during the experimental period. This difference in behavior may stem from some unobserved characteristics of the experimental and control groups, from economic differences among sites (in which racial and ethnic compositions differed), from the bias of differential attrition rates, or simply from difficulties in data collection unique for these populations. These findings need to be explained or, as Hall suggests, dismissed as unreliable.

OTHER BEHAVIORAL RESPONSES

The income maintenance project sought information about responses other than work effort to the experimental treatments and to income changes in general. These responses, such as consumption and employment behavior, education, and health status, were useful not only as control variables to explain variations in labor supply, but also as a basis for interpreting the social significance of a negative income tax. These measures were reviewed and evaluated for the conference by Peter H. Rossi, the lone sociologist among the writers.

As already indicated, the experimental families made larger investments in housing and durable goods than the control families. There also was evidence that less job turnover occurred under the more generous plans, and that turnover which did occur was mainly among younger workers shifting to better jobs. It was reported at the conference that analysis now in progress will show significant increases in educational attainment among experimental family members. But the data for health status, family composition, and individual well-being revealed few consistent and interpretable patterns among experimental families.

According to Rossi, these negative findings may reflect poor design and analysis; but to some extent, they may be the result also of the dominance of economists in the decisions on experimental design and analytical priorities. He points out, for example, that the relatively homogeneous nature of the sample population, which was considered an advantage from the economists' standpoint, made it virtually impossible to detect significant experimentally induced changes in such variables as family composition and individual well-being. He also

pointed out that there has been a persistent absence of analysis that would systematically explain economic effects in social psychological and sociological terms. In particular, the small but significant reduction in hours worked and the changes in hours of job search, in job quality, and in worker satisfaction have begged for analysis, little of which has yet been done.

The conferees did not dispute Rossi's observations regarding the experimental design, nor was there disagreement that further analysis of the nonlabor supply response would be useful. But several conferees challenged Rossi's contention that the experiment was dominated by economists. Other social scientists were heavily involved from the beginning and substantial resources were devoted to the noneconomic questions. On the other hand, some believed that it was right to emphasize the labor supply response as the dominant focus of the experiment—if only because the state of the experimental art was rudimentary at the time the experiment began.

In any event, given recent advances in experimentation and analysis, a much more productive combination of the disciplinary techniques ought to be possible in future experiments. Such a combination of talent could help to develop more of the richness of individual responses to experimental treatments; and the number and variety of the observations needed to obtain statistically significant results for a broad range of issues would require a much larger sample than was used in the New Jersey experiment.[6]

POLICY IMPLICATIONS

The policy implications of the New Jersey experiment doubtless will be debated for a long time. Opinions will differ, not only because of doubts about the methodological problems that are reviewed in this volume, but also because of the changes that have occurred in U.S. social policy during the last few years. Many advances have been made in related areas of labor-market research since the mid-1960s, and the numerous changes in income support programs—reforms in AFDC itself, the creation of supplemental security income for the aged, and the expansion of the food stamp program—have created a very different environment for income maintenance policies.

Nevertheless, many of the conferees feel that the experiment will have a significant effect on attitudes toward negative income taxation. As Rees and Watts point out in their paper in this volume: "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." Michael Barth, Larry Orr, and John Palmer note that the finding that the labor-supply response of male family heads usually takes the form of a reduction in hours of work rather than complete withdrawal from the labor force should moderate the pressure for a strict work test and shift more attention to distributive equity in decisions about guarantees and tax rates. They also emphasize that the experiment demonstrated the administrative feasibility of a negative income tax, including practical solutions for such problems as the definition of income and the establishing of accounting periods for

[6] Robert Hall contended that the labor supply results could have been replicated with a much smaller sample (see his paper in this volume), but he did not address himself to the question of sample size for determining other types of responses.

determining eligibility. Various anecdotes concerning the impact already observed in executive and legislative deliberations were cited by them and others at the conference as confirming these claims of policy relevance.

Others are skeptical about these claims. Bette and Michael Mahoney argue in their paper that the experiment's design was biased toward moderate tax rates in the belief that such rates would mean greater work effort. This bias was compounded by the absence (for analysis) of several guarantees at the 70 percent tax rate. But, the Mahoneys pointed out, the major effect of a moderate tax rate is to extend the work disincentive farther up the income distribution. They also believe that, because the price of moderate tax rates is a lessened alleviation of poverty, the emphasis on them is unfortunate.

Other reservations were expressed. The temporary nature of the experiment, combined with the intrusions of New Jersey welfare law and some of the decisions on sample composition and on guarantees and tax rates, introduces considerable uncertainty regarding the implications of the experiment for a permanent negative income tax. Moreover, many opponents of the negative income tax did not expect widespread withdrawal of workers from the labor market. The withdrawal of only a relatively few able-bodied workers was sufficient reason for much of the opposition—and the experiment did little to remove these fears. Finally, the experiment's failure to improve the accuracy of cost estimates for alternative negative income tax plans also is unfortunate.

The conferees developed little consensus on these issues. The considerations leading to moderate tax rates were vigorously defended, as was the view that work effort is discouraged by high tax rates. It was argued that equity and politics, as well as the growth in cumulative tax rates among government cash and in-kind transfer programs, suggested that a negative income tax was still the leading practical alternative. On the other hand, it was pointed out that the recent expansion of these same income support and in-kind transfer programs has greatly reduced the urgency of a comprehensive negative income tax and thus the significance of the experiment's findings.

SIGNIFICANCE OF THE EXPERIMENT

What, then, is the significance of the New Jersey income maintenance experiment for policy and for the idea of social experimentation? Eight million dollars were spent, about two-thirds of it in research costs. Aside from the direct financial costs, an enormous amount of time was spent by social scientists inside and outside the government in helping to design the experiment and in analyzing the results. Are the benefits worth the costs?

Most of the participants at the Brookings conference felt that the answer to this question is affirmative. It was generally agreed that the New Jersey experiment was conducted with diligence and intelligence, that the insights gained in program design have improved subsequent experiments, and that the administrative feasibility of a negative income tax has been demonstrated. In addition, the experiment corroborates and improves on other contemporary findings about the labor-supply response of workers to a negative income tax. As Barth, Orr, and Palmer acknowledge, however: "The experiment does not

and should not decide for policymakers whether to extend cash assistance to the working poor or at what levels and with what benefit-reduction rates. Indeed, no empirical evidence could do so. Research, no matter how relevant and competent, cannot tell us what national policy ought to be. It can provide some hard data as one input to the process that balances competing demands for scarce public resources." Moreover, even with all the demonstrable advances it has made, the New Jersey experiment remains vulnerable to those who distrust its scientific underpinnings or prefer to disregard its policy implications.

Hindsight suggested to the conferees that subsequent social experiments could improve on the New Jersey experience in several respects. Treatments should be sufficiently different so that statistically significant results will be more likely to be obtained. For example, the failure to detect the relative effects of different guarantees and tax rates was one of the major disappointments of the experiment. For the same reason, sample sizes and selection should not be too parsimonious, given the uncontrolled turbulence of the environment and the likelihood of unpredictable attrition. Pre- and post-experimental data collection opportunities should be explored more fully. Finally, the design and analysis of results should have an interdisciplinary character—for, as one conferee put it, "in no other way can we hope to explain the richness of the real world." It seems clear that any major social policy experiment will face the same type of design trade-offs and constraints that were confronted in the New Jersey trial. Yet it is impossible to say whether the state of the art in social experimentation is advancing rapidly enough—and the tempo of social change is steady enough—for the larger, more complicated experiments now under way to improve on the performance of the New Jersey experiment.

Beyond its actual and potential contributions to research in the social sciences, the New Jersey experience doubtless will have a substantial impact on the ways in which proposals for social reform are considered. Without suggesting that social progress must await the results of research, the experiment demonstrates that a new social idea need not be adopted before its consequences are appraised on the basis of a carefully controlled field test.

Policy Implications: A Skeptical View

(By Bette S. Mahoney, W. Michael Mahoney)

The New Jersey-Pennsylvania income maintenance experiment was conceived, designed, and conducted by people who tend to favor a negative income tax. This is not to say that the design, conduct, or analysis was in any way tailored to favor such a tax. The experiment was conducted with a remarkable degree of objectivity and there is no doubt that, had the results not confirmed prior expectations, they would have been treated in much the same manner as in fact the favorable results have been.

The two principal objectives of the experiment were to explore the issues related to labor supply that are posed by a negative income tax and to demonstrate scientifically that a negative income tax would not produce a large-scale reduction in the labor force. The direct policy implications of the experiment relate to issues of work incentives and the question of the probable cost of a negative income tax. But

although the importance of the labor supply effects of a negative income tax has been stressed repeatedly, little attention has been paid to the way in which such knowledge could be incorporated into policy. The first section of this paper is an attempt to remedy this deficiency. The second is a discussion of what the experiment tells us about the design of a negative income tax program.

No single experiment or research effort could hope to address all issues relevant to income maintenance policy. But the concentration on one aspect of the income maintenance policy debate—labor supply—will limit the impact of the experiment. In this context, the third section of this paper examines the New Jersey experiment as an example of research in social policy.

THE OPTIMAL TAX RATE

From the time a negative income tax was first seriously proposed, the appropriate marginal tax rate has been considered a critical feature. The lower the rate, the greater the number who will be eligible for benefits and, other things constant, the greater will be the cost. The lower the rate, however, the more also financial disincentives for work are minimized, and the less likely it is that higher payments will result from reduced earnings. The first part of this section discusses what the appropriate tax rate would be if only work effort and program cost are considerations; the second part, whether other considerations, especially equity, also should play a part.

Guarantees and Tax Rates, Labor Supply and Cost

Both the level of the guarantee and the marginal tax rate—that is, the tax in cents deducted from the guarantee for each dollar of additional income—may affect work effort, but the tax rates have received the most attention—from the conservative Milton Friedman and the liberal James Tobin alike. Tobin argues for minimal standards of assistance to keep families from falling below the poverty line; that "the schedule of benefits . . . must provide incentives to work" is another and separate principle.[1] Friedman says about the marginal tax rate that "fifty percent is too high. I should prefer less."[2] Throughout the professional literature on income maintenance policy the pattern is the same: it is through marginal tax rates that the issue of work disincentive is addressed.[3] In contrast, most of the concern of politicians has been over the incentive effect of guarantees.

From the mid-1960s to date, proponents of the negative income tax have stressed the importance of low tax rates by highlighting the presumed 100 percent rate in existing welfare programs. They also stressed the universal coverage of their proposal by highlighting the presumed exclusion from assistance of the working poor. In fact, neither of these views of the existing welfare system was completely

---

[1] James Tobin, "First Lecture," *Welfare Programs: An Economic Appraisal*, Rational Debate Seminar Series (American Enterprise Institute for Public Policy Research, 1968), p. 16.

[2] Milton Friedman, "The Case for a Negative Income Tax: A View from the Right" (paper prepared for the National Symposium on Guaranteed Income, Chamber of Commerce of the United States, December 1966; processed). Friedman nonetheless opted for a 50 percent rate because, in the logic of his form of negative income tax, lower rates would have meant inadequate benefits for families without income.

[3] This is a reference to the policy discussions; the New Jersey experiment had greater variation in guarantee levels than in marginal tax rates.

correct. Nonetheless, the hallmark features of the negative income tax came to be that it provides for a marginal tax rate of less than 100 percent and for coverage of the working poor.

The perceived need to have low marginal tax rates is evident in the design of the New Jersey experiment: no treatment group was subject to a marginal rate higher than 70 percent, and in determining the number of recipients to be assigned to the various plans, it was assumed that policy interest in 50 percent rates was more than three times greater than in 70 percent rates.[4] Thus, the design of an experiment to determine the labor supply response to different marginal tax rates incorporated the assumption that a 50 percent rate was likely to be preferable from the standpoint of social policy. The basis for this assumption is not clear.

The problem of the tax rate has been posed by many critics as one of providing work incentives. Work incentives would be as cheaply provided by a low guarantee or by having no program at all. A program that gives people income for which they do not work necessarily discourages work and does not encourage it. Thus, the real problem is how to mitigate or offset the disincentives that are inherent in the very nature of assistance programs. Further, what we are really concerned with is not disincentives or incentives as such but the behavior they induce—in this case, work. Whether or not the economic disincentives present in varying degrees in different income maintenance plans are offset by psychological, sociological, or other incentives is an empirical question that the experiment, the cross-sectional studies, and other research have attempted to answer. The constant references to the need for low tax rates have obscured this important point.

But why all the concern about work? Beyond the fact that this is a major concern of politicians, there are two main reasons. The first is because of an interest in program cost; the second is because of the common view that it is better for most people to work than not to.

EFFICIENCY AND COST. A basic objective of income transfer programs for workers is to raise their income, not relieve the necessity for work. If guaranteeing a worker a minimal income means that he will quit and be no better off, why bother? But when the program covers a large group of people, some of whom are expected to work and some not, but between whom there is no easy dividing line, the problem is more complicated. The possibility of paying somebody more than he deserves—in the sense that the person should be working and receiving less subsidy—is unavoidable. If workers do quit or reduce their work effort, then either the program will cost more than otherwise would be the case or the guarantee will be less.

Because funds available for alleviating poverty or for other public purposes are limited, it is appropriate to consider ways of reducing unintended side effects that make the program less efficient than it could be. Relatively low marginal tax rates have been the generally recommended method of keeping workers from quitting. If the objective of the program is simply to raise the incomes of the poor, then the marginal tax rate should be set at the level which for a given cost would produce the highest possible guarantee. This objective may not imply lower rates. Lowering marginal tax rates might not produce sufficiently

[4] See the paper by Felicity Skidmore, "Operational Design of the Experiment," in this volume.

greater work effort to make transfer costs go down: for this to happen, work effort has to be fairly sensitive to the tax rate. In this sense, efficiency in a negative income tax is more likely to require a high rather than a low rate; and to determine the optimal rate would require knowledge about both the income distribution and the labor supply response.

WORK AS AN INTRINSIC GOOD. Though efficiency criteria imply high marginal tax rates—even as high as 100 percent—a desire to maximize work effort may prompt lower ones.

Increases and decreases in the gross national product traditionally have been regarded as net gains or losses for society at large: if someone works, it is good for us all. But for the individual taxpayer, the total national product is not of interest, only his share in it. Those who must pay taxes so that the earnings of others may be supplemented have the right to ask what they gain from any work that is thereby induced. If the individual whose income is supplemented works more but also consumes more, no direct benefit may accrue to the taxpayer.

A fear also may exist that the supply of low-wage, unskilled workers and the relatively cheap goods and services they produce will be diminished. To the extent that this fear actually is realized, a true economic loss occurs. The extent of the loss is difficult to measure, however, because people's consumption patterns will change as prices and income change. On the other hand, to the extent that people with higher incomes pay more to employ the lowest paid under a negative income tax, the income distribution will change in a progressive fashion.

It may be that work effort is viewed as a good in its own right for a complex set of reasons, including the puritan ethic and other philosophies of life that hold idleness to be immoral and offensive. But society has not found idleness so offensive that it discourages it with methods beyond economic sanctions and peer-group pressure. It may be that society is willing to accept the degree of idleness that now prevails—but would be unwilling to accept more.

Moral and ethical concern about people who seek to live at the expense of others focuses primarily on a dichotomy of work versus no work. Programs designed to induce more work from those already working might not be considered as attractive as those which produce more workers. Few persons would argue for expensive incentive features solely on the basis that less expensive schemes would discourage overtime and dual job holding. This means that in examining the sensitivity of labor supply to marginal tax rates and guarantee levels, it is essential, first, to distinguish between outright entries into, or withdrawals from, the labor force and mere additions or reductions in effort; and, second, to specify precisely what kinds of additional work effort are being sought.

This issue is even more complex when the question of secondary workers, part-time workers, and length of job searches is considered. It seems unlikely that many persons would wish to commit funds for the sole purpose of inducing more work effort from secondary workers.

In brief, we do not know how or even whether work as an intrinsic good is preferred by a majority in society. Society clearly should not spend beyond the level at which no additional work is induced; paying for expensive incentive features that do not affect behavior makes no sense. That lower marginal tax rates will bring about a greater total work effort is not at all clear. Modest reductions in the marginal rate faced by people below the original breakeven point will require larger

increases in the rate for the newly eligible above it, who previously had been subject only to the positive income tax. Suppose that it is proposed that 80 percent tax rates in a plan with a $4,000 guarantee be reduced to 50 percent. This would represent a 38 percent reduction in the rate faced by those originally eligible. But if the rate above the breakeven point had been 20 percent, there would be a 150 percent increase for the newly eligible.

Marginal Tax Rates and Tax Equity

In recent years, two reasons unrelated to labor supply have been suggested for preferring low tax rates in a negative income tax: first, that high rates conflict with the requirements of vertical equity; and, second, that it is unfair to tax the earnings of the poor. Both are intuitively appealing arguments; both are wrong.

The concept of vertical equity developed out of taxation theory and is generally expressed as "those who earn more should end up with more." But there is nothing in this concept that serves as a guide to how much more. Thus, a marginal tax rate of 99 percent would satisfy the requirements of vertical equity. The concept is more correctly expressed as "those who earn more should not end up with less." The word "equity" connotes fairness and justice, concepts that are deeply embedded in the fabric of our society but that nonetheless are ambiguous.[5]

The meaning of the term "tax rate" also has become troublesome during the debate over welfare reform. The term can refer to average or marginal taxes. In the context of a negative income tax, families have usually been viewed as paying a positive tax if they were above the breakeven point, nothing if they were at that point, and a negative tax if they were below. This view refers to average taxes; the family faces positive marginal tax rates below, at, and above the breakeven point, rates that were regarded as high below the breakeven and low above it. We believe more attention should be given to average tax rates, which show who pays and who receives and thus provide a more comprehensive picture of the nation's tax and transfer system. This also would clarify certain of the equity issues associated with tax rates in the welfare reform debate.

In this debate, it has been argued that only the rich and the poor face high marginal tax rates and that this is unfair to the poor. But in terms of average tax rates, the poor are net recipients of transfers and their average tax rate is negative, which by definition is not high. People who complain that the poor face high marginal tax rates under a benefit program are, in effect, taking the program's guarantee for granted. When a recipient works, the welfare benefits he forgoes as his payments are reduced are tantamount to taxes paid. But all families, not just those receiving reduced benefits, must be considered to have forgone welfare benefits—in effect, to have been subject to a tax. In this view, families from just above the breakeven point to quite far up the scale of income distribution also face high average tax rates.[6]

---

[5] See Martin Bronfenbrenner, "Equality and Equity," *Annals of the American Academy of Political and Social Science*, vol. 409 (September 1973), pp. 9-23; and John Rawls, *A Theory of Justice* (Harvard University Press, 1971).

[6] Consider a negative income tax with a $4,000 guarantee and a 50-percent marginal tax rate. At the breakeven point of $8,000, the net additional disposable income over and above the guarantee is only $4,000. Similiarly, if the marginal rate of taxation were 20 percent, a family with $10,000 of earnings will have only $6,000 more disposable income *less* the $400 it would pay in positive taxes. Thus, in terms of disposable income, a family with $10,000 of earnings has only $5,600 more than another family with no earnings. On this basis, the $10,000 family may be regarded as having been subject to a 44-perecnt average tax rate.

The problem of taxing the earnings of the poor is a sensitive matter, it being perhaps the mark of civilized society that income taxes are not to be imposed on the destitute. Surely in our wealthy society there is no need to require the poor to share in the cost of public programs. It is a virtue of an income tax over other forms of taxes that the poor can be exempted. But there is no similar virtue in a negative income tax: a positive marginal tax is a necessary condition thereto. In fact, high rates concentrate expenditures on the poorest of the poor, whereas low rates mean that funds will go to the less poor—sometimes to those who clearly are not poor. If fairness is measured by the proportion of the expenditures received by the poor, high tax rates are fair, and low tax rates may be viewed as part of a policy to alter the shape of income distribution considerably beyond the poverty level.

If there are no arguments in equity or in efficiency for low tax rates, are there any others? It has been suggested that marginal tax rates may cause such other behavior as welfare cheating or marital instability. If society wants more or less of such behavior, that fact certainly can be considered in setting marginal tax rates, provided the incentive has some empirical, as well as theoretical, substance.

IMPLICATIONS FOR THE DESIGN OF A NEGATIVE INCOME TAX

The New Jersey experiment was a partial simulation of a negative income tax. From its results can be drawn inferences about the appropriate tax rate for a negative income tax. Inferences also can be drawn about other aspects of income maintenance policy. There are some things the experiment can tell us about welfare policy and some that it cannot.

The Marginal Tax Rate

Suppose that the information generated by the New Jersey experiment provides the basis for the design of a negative income tax. Setting aside the need for information on other issues and the important question of whether the results truly reflect the impact of a full-scale, permanent program, the New Jersey results can be summarized in this manner: A negative income tax with guarantees at or near the poverty level and marginal tax rates between 30 percent and 50 percent will cause a reduction in the amount of labor supplied by low-income households of about 5 percent, certainly no more than 10 percent. Within the stipulated range, differences in the marginal tax rate do not cause different labor supply response. Secondary workers will reduce their work effort more than primary workers. Overall, reductions in work effort do not come from outright withdrawal from the labor force but through reduced hours of work, dual job holding, and to some extent slightly longer periods of unemployment.

In line with the arguments of the preceding section, these results clearly suggest that the marginal tax rate should be at least 50 percent—presuming, of course, that the income of eligibles would not be subject to other taxes. This conclusion can be reached without reference to income distribution data for both objectives of efficiency and work inducement; it follows from the finding that different tax rates produce no differences in labor supply.

Could the tax rate be higher than 50 percent? Unfortunately, because of contamination of the 70 percent treatment by other welfare programs and failure to include treatments at even higher tax levels, the experiment provides little basis for inferences about the labor sup-

ply responses to rates higher than 50 percent. Nonetheless, to oppose a higher tax rate, one would have to believe that it would cause a sharp increase in labor supply withdrawal.

Tax rates higher than 50 percent might produce some cost savings and some further reductions in work effort among eligibles, although perhaps not a net reduction in work effort. In any event, so long as additional reductions were to manifest themselves as modest reductions in hours worked and dual job holding and not in withdrawal from the labor force by primary workers, they probably would not be considered socially undesirable.

The Work Requirement

In the traditional form of a negative income tax there would be no such thing as a work requirement: that is, benefits would depend solely on income and not on work status. This controversial feature of the negative income tax is deemed essential by proponents but provides the basis for opposition from others. The experiment was not designed to illuminate discussions of the feasibility or desirability of a work requirement. Nonetheless, its findings that the poor have relatively strong labor force attachment, like similar findings in the past, have been used to support the view that work requirements are not necessary, thus perpetuating a misconception of the role and possible effectiveness of a work requirement.[7]

Surely, if an income assistance program produced large-scale withdrawal from the labor force and efforts to beat the system were widespread, no army of bureaucrats could enforce a work requirement. If the reverse were true—that is, if there were few efforts to get away with something—a work requirement very well could be designed and administered in a manner that was equitable and firm and preserved public faith in the program. The New Jersey results strongly suggest this possibility.

To maintain program integrity, a work requirement may be considered essential regardless of the rate of marginal tax. A work requirement also may offset some disincentive effects of higher marginal tax rates. If the real policy alternatives lay between, first, a modest guarantee, a relatively low tax rate, and no work requirement and, second, a generous guarantee, a high tax rate, and a work requirement, the not-so-poor, who would receive income supplements with a low tax rate, might side with those who oppose the work requirement. But it seems likely that the very poor would agree with the general public and endorse a work requirement.[8]

Compound Marginal Tax Rates

Much has been written recently about the high marginal tax rates that can result when several programs condition their benefits on income. There is no reason to assume, however, that the disincentive effects of an all-cash program would be the same as a combination of cash and in-kind programs, even where the apparent marginal rates are the same. There is, in fact, some reason to believe that the disincentives from a combination program might be less. For example, if a family's income increases to the point at which the family felt it could afford better housing than is available in public housing programs,

---

[7] See U.S. Department of Health, Education, and Welfare, "Summary Report: New Jersey Graduated Work Incentive Experiemnt" (1973; processed), p. 45.

[8] The level of guarantee constitutes the actual work requirement. An administrative work requirement is only a device to lower or eliminate the guarantee when an ablebodied individual refuses to work.

it might disregard the loss in housing benefits. An individual going to work where he would participate in an employer-subsidized health insurance program may substantially discount his loss of medicaid benefits.

Enough of this kind of reasoning can be developed to suggest that the very attempt to calculate arithmetically the compound rate of in-cash and in-kind programs is futile. The differing participation rates associated with in-kind programs also are important. "Selection-out" by participants at higher levels of income generally is thought to be more commonly associated with in-kind than with in-cash programs. If this is true, programs otherwise similarly designed will differ in the degree to which their benefits are distributed to the poorest households.

Discussions of the advantages of cash benefits over in-kind benefits have focused on maximizing utility to the recipients and, more recently, on maximizing the utility to both taxpayers and recipients taken together. The political attractiveness of in-kind programs is well known, of course, and many have acknowledged that they would prefer having in-kind programs to assist the poor rather than having no program at all. The point here is somewhat different: in-kind benefits may provide greater total benefits to the poorest households because the nonparticipation of those with higher incomes will allow more generous benefits to the poor. This important area of debate is one to which the experiment contributes little. It is, in fact, an area that has become much more important since the experiment was launched.

Cost of a Negative Income Tax

A third and final topic that the experiment tells us little about is the increase in transfer cost that would be associated with the adoption of a negative income tax. This is the result in part of the existence of a nearly universal food stamp program that already may have caused some of the labor supply effects that would be associated with a negative income tax, and in part of the ambiguity of the results. In addition, a nationwide negative income tax would include numerous groups that were not eligible under the experiment. Moreover, shifting from a food stamp program to a cash program might increase participation, but the results of the New Jersey experiment cannot be used to estimate the extent of the increase.

THE NEW JERSEY EXPERIMENT AS POLICY RESEARCH

At the time it was undertaken, the New Jersey experiment was among the government's most ambitious and expensive projects in social policy research. With the advantage of hindsight, the experiment now provides us some important information about experiments in particular and about policy research in general. The following paragraphs contain a discussion of some of what has been learned.

Some Lessons for Planning

In their overview of the New Jersey results, Rees and Watts say: "If there were people who expected our experimental treatment to cause large declines in the [labor force] participation rates of male heads of households, they were not in our research group." [*] It seems likely that neither were they in the HEW or OEO offices sponsoring the experiment. Why, then, was an expensive program of experimentation in income maintenance programs undertaken?

[*] See Albert Rees and Harold W. Watts, "An Overview of the Labor Supply Results," in this volume.

First, those persons who were urging the enactment of a negative income tax felt that it was necessary to obtain better proof that the labor force effects would be acceptable; but "how much" would be considered acceptable was never defined.[10] Moreover, whatever the anticipated levels, they were expected to vary from one marginal tax rate to another—but just how they would vary was not set forth. Why 100, 90, and 80 percent marginal tax rates were expected to produce unacceptable levels of withdrawal whereas a 50 percent rate would not is unclear.

If the arguments of the first section have any merit, relatively high marginal tax rates might be preferable to low rates. Both by design and by accident, the New Jersey experiment provides little information about the impact of higher tax rates. The impact of the experiment may be limited because policymakers may choose to defer decisions until such information does become available. Certainly, one requirement for an effective experiment must be a thorough examination of all the policy alternatives and issues.

Second, in the view of its advocates, the proof that disincentive effects of a negative income tax were small had to be rigorous. This implied careful use of controls rather than the less formal and less expensive demonstrations and analyses of program data. Moreover, the results had to be quantitative: it was essential to measure the disincentive effects as well as to make a judgment as to whether they were large or small. Finally, the results had to show the impact of variations in guarantees and tax rate.

The failure to exploit the research potential inherent in existing programs is in part attributable to this preference for the "rigorous proof of experimentation" and in part of ignorance about such programs. To our knowledge, neither the AFDC-UP program nor the state-run working-poor programs—some of which have 100 percent marginal tax rates and some of which do not—have received anything remotely approaching the investment of funds, talent, and effort that have been expended on the experiments and cross-sectional analyses in the negative income tax program. Ignorance about these programs also may have contributed to undue concern about labor supply response. Much of the information about how existing programs actually operate, which the experiment was to reveal, was and is readily available without resorting to experimentation. Certainly, before experimentation is undertaken, other potential sources of information should be fully explored and analyzed.

As to the rigor of experimentation, rigor may have been unattainable or unnecessary. Questions about the feasibility of simulating a full-scale, permanent national program with a short-term, random, local experiment were raised from the very beginning. These questions never were answered and have been reinforced by the ambiguities and paradoxes of the experimental results. Moreover, if small declines in labor supply are expected, small differences among different treatments also must be expected. For policy purposes, once it is known that the differences will be small, there is no great value in knowing exactly how small. For firm evidence that the decline would be small, analysis of existing programs, coupled with a set of demonstrations, could have been performed more quickly and less expensively.

---

[10] It would be ironic if those who opposed the negative income tax because of fear about work effort actually expected the same amount of labor force withdrawal as those who supported it.

Third, recent research suggests that the costs of a negative income tax are not sensitive to changes in labor supply.[11] At least for the types of plans that generally have been advocated, greater precision in cost estimates than is offered by the experiment could be obtained by improving data sources and exploring other factors affecting cost, such as the propensity of program eligibles to participate. For example, the supplementary security income program has had fewer applicants than were estimated: either its data bases, the participation rate assumptions, or both were in error.

The Contribution of the New Jersey Experiment

The New Jersey experiment already has had an impact. It has made experimentation in the social sciences respectable and has helped spawn not only three other income maintenance experiments but also an elaborate program of experiments in housing allowances and in health insurance. It has had other impact as well. During the first round of hearings on the family assistance plan, the House Ways and Means Committee invited Harold Watts, Lee Bawden, and David Kershaw to testify in executive session. One of the authors was present throughout the hearings, and it seemed to him that their testimony was an invaluable contribution: a program similar to the family assistance plan, although on a much smaller scale, actually was being run, and no abrupt withdrawals from the labor force had been observed; possibly the concern about work incentive was misplaced.

The experiment demonstrated another value of experimenting before implementing: the opportunity to think through the multitude of administrative and definitional problems associated with turning an idea into an actuality. The experience of the New Jersey experiment, augmented by the experience of the states in operating existing welfare programs, will furnish the basis for the administrative features of any new program.

What about the future impact? Unfortunately, the program of experiments concentrated on negative income tax programs to the virtual exclusion of alternative approaches to income maintenance. Therefore, although we now have a great deal more information about the operation and labor supply effects of a negative income tax, we have little additional information about alternatives. Proponents of programs as diverse as wage subsidies, guaranteed employment, family or children's allowances, and expanded social insurance may feel that it would be appropriate to have more information available about such proposals before this country selects its course on an income maintenance policy. This fact alone will temper the impact of the experiment.

Moreover, additional information is needed about negative income tax programs, including data regarding:

—The employability of female heads of families
—The design and potential effectiveness of work requirements
—The programmatic impact and equity implications of alternative accounting periods
—The effect of compound marginal tax rates, especially where they result from in-kind transfer programs
—The effect of taxing income aside from wages and salaries
—Ways to vary guarantee levels by family size

[11] Ralph D. Husby, "Work Incentives and the Cost Effectiveness of Income Maintenance Programs," *Quarterly Review of Economics and Business*, vol. 13 (Spring 1973), pp. 7–13.

—How and whether to supplement a negative income tax or its alternatives with a program to meet emergency and special needs

What these additional needs clearly suggest is that for a social experiment to have a truly significant impact on the policy development process, it must be part of comprehensive and balanced programs of research and analysis.

But in terms of what it set out to do—to illuminate labor supply effects—what has the experiment accomplished? That cash assistance will encourage the poor not to work is a deep-seated fear, a depressingly constant theme of discussions of welfare policy stretching back beyond Elizabethan poor law. For all practical purposes, the statement that the able-bodied poor should not be given welfare can be as much a moral dictum as it is a corollary to an economic prediction. It seems unlikely that the New Jersey results will still such fears: too many questions can be raised about the validity of the experiment. In 1968, Guy Orcutt and Alice Orcutt suggested that the successes of experimentation in the biological and physical sciences would lead the public to accept experimentation in the social sciencies.[12] But accepting experimentation and accepting results as reported by the experimenters are not the same thing. Those persons who originally believed that there would be small disincentive effects may well accept the New Jersey results, whereas many of those who believed otherwise may well remain unconverted.

CONCLUSION

The New Jersey experiment was undertaken at a time when many economists assumed a priori that low marginal tax rates were preferable to high and had favorable implications for labor supply. With respect to program cost, however, high tax rates may be preferable. Moreover, low tax rates do not lead unambiguously to increases in labor supply. But unfortunately, the design of the experiment was dictated by the assumption of the times.

The experiment was under taken to prove that a negative income tax would not induce large-scale withdrawal from the labor force; to measure the difference in responses to different combinations of guarantees and marginal tax rates; and to provide the basis for estimating the cost of a national program. But for purposes of policy formulation, existing programs or demonstrations could be just as suitable for stilling unwarranted fears about labor supply; small differences in response do not have important policy implications; and other factors contribute more to uncertainty about program cost than does labor supply.

Experiments in social policy should be considered in relation to their cost and to the likelihood that some spectrum of opinion will be shifted—and how far. In our view the New Jersey experiment does not stand up well to these measures. It was conducted with intelligence, perseverance, objectivity, and wit, and yet it produced something less than overwhelming evidence.

Sunk costs are sunk. But it is not too late to consider whether the continuation or expansion of the current experiments is worthwhile. The cost of all the income maintenance experiments is approaching $70

[12] Guy H. Orcutt and Alice G. Orcutt, "Incentive and Disincentive Experimentation for Income Maintenance Policy Purposes," *American Economic Review*, vol. 58 (September 1968), pp. 754–72.

million: the many permutations of labor supply response to negative income taxes that these experiments will test are not worth the cost. Although ethics require that promises to existing participants be kept, expansion to explore other issues should be seriously questioned.

None of the foregoing should be construed as implying that the New Jersey experiment was a failure or a waste of time and money. Among other things, it was an experiment in experimentation. The experiment produced extensive data about labor force behavior at a cost of $8 million, which seems quite reasonable when compared with the costs of such other research information as the Survey of Economic Opportunity of the U.S. Bureau of the Census or the Panel Study of Income Dynamics, the longitudinal survey of the Survey Research Center of the University of Michigan. It also produced information about the design and administration of welfare programs, which, although not an argument for experimentation, is valuable nonetheless. Finally, the experiment did generate another and unique observation: that for relatively low levels of welfare, the poor will not opt for leisure. This, and the other observations available, eventually may convince the doubters.

COMMENT BY RICHARD P. NATHAN

As a practitioner in welfare policymaking, I was asked to comment on the Mahoneys' paper. I was a member of that group of welfare planners who, in Gilbert Steiner's words, went blithely ahead on the family assistance plan in 1969 despite the fact that the New Jersey experiment was well under way and presumably would produce findings pertinent to these decisions.

The Mahoneys make an important contribution by their thoughtful treatment of the principal question of the New Jersey experiment: namely, the role of the marginal tax rate in income support programs. We learned such lessons the hard way in designing the family assistance plan: the lower the marginal tax rate, the more resources are shifted from the poor to the near poor or nonpoor. If insufficient resources are available to allow income support programs to eliminate poverty fully, the issue becomes how much poverty should be retained in the interest of maintaining strong work incentives.

I was pleased also that the Mahoneys included in their paper a discussion of the role and efficiency of in-kind transfers. As a proponent of a strategy of incremental reform for income support programs at this time, I believe we should devote more attention to this subject. The food stamp program now provides benefits to most poor working families at a higher level than would have been the case under the family assistance plan.

I want to distinguish between design problems in the New Jersey experiment and what I shall call questions of policy relevance. Henry Aaron's paper provides the framework for looking at this distinction. Aaron lists what he calls the "acknowledged problems" of the New Jersey experiment. The first two of these—the brevity of the experiment and the introduction of an AFDC–UP program—are design problems that make extrapolation of the results hazardous; both presumably could be corrected in the design and implementation of future or current experiments.

But Aaron's third acknowledged problem gets at what I define as the larger question of policy relevance. He states that "the thinness of the sample and the brevity of the experiment make it impossible to

observe the impact of a negative income tax on the *mores* of entire groups" (emphasis added). The use of the word "mores" raises for me the question of whether the kind of a prolonged and emotional debate that would be necessary to pass a negative income tax would result in changes in behavior in terms of the choices between work and leisure that will be made by eligible or potentially eligible persons. Put another way, even if all the design problems of the New Jersey experiment were resolved—and this is no easy task—the basic question remains as to whether the findings from such an experiment would apply after a national and highly visible shift in policy—such as adoption of a negative income tax would represent—had occurred. Bluntly stated, is it not possible that adoption of a so-called "guaranteed-income" program would be interpreted by the eligible population as a congressional sanction for leisure?

What are the answers to this criticism of the New Jersey experiment in terms of its relevance for national policymaking?

One answer is that in Seattle an attempt is being made to advertise the availability of the new income maintenance program. This, it is suggested, will allow the researcher to claim that people understood the policy change and therefore that the experiment has taken into account the full impact of the new program on public opinion. I do not consider this an adequate answer to my suggestion that a change in mores could take place after a negative income tax had been enacted. We are not interested in how people's attitudes change in one locale in response to a program of limited duration. We are interested in how attitudes would change nationally after a long and emotional debate—as part of the broader debate over the adoption of a nationwide negative income tax program—has taken place on the fundamental question of the employment obligations of the poor.

A second answer is simply that this line of criticism could be directed against all social experimentation and that I am overstating its importance. Again, I am not satisfied. If we assume limited resources for social experimentation, as we must, my argument is that we should select areas for experimentation in which our findings are most likely to be relevant to policy and used accordingly. We could experiment in many social program areas. Generally, however, we should hold off experimenting in areas in which policy change is apt to be of so fundamental a nature and so emotionally charged that concern must necessarily exist about the effect on behavior of the policy changes being studied—and hence about whether the findings will remain relevant once the policy has been adopted and put into operation. Experiments in such areas are less useful than those which we judge will not involve policy changes likely to affect behavior on a broad scale: for example, new types of manpower, child-care, educational, and health-financing programs.

I conclude, therefore, that at the very least we must give more attention to efficient use of the resources available for social experimentation, and that we must develop strategies that weigh the potential payoffs of different types of social experimentation.

COMMENT BY ALAIR A. TOWNSEND AND JAMES R. STOREY

We applaud the Mahoneys for reminding us that tax rates are only one parameter in the highly complex structure of an income maintenance system. Analysts probably have overindulged themselves in

the study of tax rates, perhaps because they prefer to study problems that are well structured and quantitative and to which high-powered analytical tools can be applied. In addition, many of the analysts are economists, who naturally focus on policy issues related to labor markets.

The Mahoney paper, however, overreacts to the tax-rate focus. The tax rate is, after all, the single most significant factor in most income maintenance programs. It not only affects work incentives and program costs but also is a primary determinant of income adequacy for recipients with income—as well as of equity, both among recipients and between recipients and nonrecipients. Because the tax rate represents the marginal response of the system to any marginal change affecting income, the rate level and structure are crucial in many ways. ways.

But although analysts may have been preoccupied with tax rates, politicians have not—at least not in the broad context just outlined. As a result, many programs of the negative income tax type are operating at all levels of government, with tax rates ranging from 25 percent to 100 percent—or even higher if one includes notches. The present conglomeration of programs includes such bizarre contrasts as states applying 100 percent tax rates to working men in general assistance programs, while the federal government taxes at 50 percent the wages of aged, blind, and disabled recipients of supplemental security income and social security.

The Mahoneys suggest that, in the compromise among adequacy, costs, coverage, work incentives, and equity, the greatest weight should be given to raising guarantees while minimizing costs. On these grounds, many current programs come remarkable close to meeting their objectives. We disagree that the balance should be struck in this way. Our view of the problems to be addressed by income maintenance reform leads us to urge moderate guarantees and tax rates.

In our view, a negative income tax or other income-related reform plan must accomplish three financial tasks: [13]

—Higher incomes for persons with little or no income or income-producing opportunities who either are currently excluded from categorical coverage or live in states paying well below average benefits
—Higher incomes for persons with income that is inadequate
—Constraining to reasonable levels combined benefits and tax rates that result from multiple programs

Given these priority objectives, it is difficult to see how high tax rates could supplement modest wages or facilitate program coordination.

Exploring Equity

The concepts of equity applied by the Mahoneys seem to be only straw men. Extreme positions have been taken along dimensions on which there is great room for compromise.

They have taken a restrictive view of vertical equity—that it forbids only a reversal of income positions. By this definition, programs that equalize or nearly equalize unequal pretransfer income positions meet the criteria for vertical equity. The definition of vertical equity with which we are familiar and comfortable requires some degree of posttransfer differentiation among former unequals. We expect pop-

[13] The tasks noted here relate only to financial parameters. Other important targets of reform are improved administration and more standardized treatment of recipients.

ular acceptance of this view to be manifested by increasing concern over the fact that millions of social security beneficiaries who receive added income from the supplemental security income program or supplemental state programs received only $20 per month as a bonus for their social security contributions. A further result is that persons with markedly different previous earnings levels and social security benefits will receive identical posttransfer income.

Another straw-man concept of equity is that the poor should not pay taxes out of funds that were rightfully theirs. We do not know of serious analysis in which this argument has been advanced. Rather, we believe the argument is that people should not be made worse off for having worked.

It seems to us that there would be wide agreement in Congress—if the proposition were phrased in straightforward terms—that those who earn more should have more income. That the Congress does not always act this way is another matter. We believe that these considerations are only now becoming understood, and that the legislative results may change as a consequence.

The Mahoneys' willingness to level incomes with 100 percent tax rates disturbs us and no doubt would disturb many members of Congress as well. Furthermore, implicit in their own questioning of the New Jersey income maintenance experiment seems to be a thread of doubt that they can state so categorically that tax rates do not matter. Finally, it seems to us to be socially destructive to apply a 100 percent or nearly 100 percent tax rate to the very bottom end of the income distribution. Such a policy in effect says that the work efforts of recipients are meaningless.

Equity is a slippery concept, especially because it is virtually impossible to separate out equal opportunities from equal outcomes. That is, do people have low pretransfer income because they have had unequal opportunities or because they have applied unequal efforts? If in part the answer is unequal opportunity, is it fair to treat such people as if they had an equal chance? It seems virtually impossible to make such determinations. Thus, a moderate guarantee for those who have had unequal opportunities provides something in the way of recompense, and a moderate tax rate then helps to differentiate among people on the presumption of unequal efforts.

To us, the equity issue revolves in large part around the bases for benefit distribution. The primary question is how far one wants to deviate from market rewards rather than build on them in the absence of any other clear basis for distribution. In practice, decisions about whom to aid and how much to aid them become decisions about whom and what behavior to reward.

Guarantee Levels and Adequacy

The Mahoneys stress their view that the role of tax rates is apparently to keep costs low and the role of the guarantee is to give people adequate incomes. We disagree. We would state the objective of both tax rates and guarantees as the maximization of recipients' total incomes in a fair way. This focus on goals is important.

The Mahoneys neglect the role of tax rates in reducing poverty and raising incomes. High tax rates essentially prevent people from raising their incomes except through very large jumps in earnings. It should be clear that in determining total income the tax rate is as important or more important for many people than the guarantee.

Achieving income adequacy is nowhere nearly as simple as the Mahoneys present the matter. Because federal guarantees are unlikely ever to achieve levels that are accepted by everyone as adequate, helping the poor means building on their own efforts by applying a relatively low tax rate to their earnings.

Moreover, it can be argued legitimately that guarantee levels—at least in a federal program—should not be set at adequate levels, where adequacy is defined as benefits sufficiently high that supplementation from private sources is not required. First, there is the concern about the impact of such benefit levels on labor force participation. Only by ignoring the potential work-reduction effects of high guarantees can one argue wholeheartedly in favor of putting most of the dollars into the guarantee. There is disquieting research on AFDC that suggests that guarantees have a greater impact on work than do tax rates. Second, relatively few persons under sixty-five have no private income or income-producing opportunities. Most persons and families thus require income supplementation, not total income support. Therefore, to construct a federal program oriented largely to the few truly destitute seems misguided. The needs of special cases are better met under more subjectively operated state-operated supplemental programs.

The Mahoneys properly stress that the negative income tax can only reduce hours of work, not increase them. They argue that analysts and others have focused too much on the tax rate, to the exclusion of the guarantee. This statement certainly does not hold for most politicians, who seem to worry about one of two things: either the impact of giving aid to employables or how to raise benefits to high levels for groups that arouse sympathy. And with respect to the family assistance plan, the tax-rate issue was merely a convenient way to scuttle a plan that was objectionable largely because it offered an income guarantee to male-headed families. Within weeks of the plan's demise the Senate Finance Committee reported out a provision that would have raised the AFDC tax rate on working women. We find little evidence that policymakers are excessively concerned about tax rates. Indeed, there is room for much greater concern about tax rates, especially from combined programs.

Of course, one factor pushing up the guarantee level is the problem of integrating a negative income tax with existing programs. The Mahoneys' paper gives little attention to program integration, but the vested interests of old programs and the objectives of new programs dictate some kind of integration. The Mahoneys hold that to compare benefits and tax rates from combined programs is a misleading and irrelevant exercise. They quite properly point out the technical difficulties and the extreme assumptions one has to make to construct such tables. But they offer no alternative. A look at combined programs is important, whether one looks at combined benefits or combined tax rates.

The tax-free combination of food stamps and AFDC, for example, is reaching astonishing levels: in eleven states it ranges from $3,900 to $4,500 for a penniless family of four and in eleven states, from $4,500 to $5,000; in five states it exceeds $5,000 a year. In the median state the combined benefit of $4,092 is equivalent to $5,360 in gross earnings (after taxes and work expenses of 15 percent of earnings). In New York State, this gross earned income equivalent rises to $7,020. At one time, food stamp tax rates were not significant for most AFDC

recipients. In low-benefit states such as Indiana, Mississippi, and Missouri, there was virtually no AFDC tax rate over large ranges of income because of state use of maximums and other methods of computing payment amounts. In such high-benefit states as New Jersey and New York, recipients usually received the minimum food stamp bonus so long as they were eligible for AFDC. But tax-rate additivity is more important now because of increases in the food stamp allotments and breakevens.

Value of the New Jersey Experiment

We agree with many of the technical criticisms the Mahoneys level at the design of the New Jersey experiment. Much of the criticism of the experiment at the policy level clearly comes from years of further knowledge and experience accompanied by 20–20 hindsight. For instance, the Mahoneys continually stress that the experiments are limited to only one type of program. But the experiments followed on the work of a number of government task forces and agencies that analyzed the negative income tax as only one of several alternatives, from which it emerged as the vehicle of choice within government. Why should federal funds have been used to experiment with ideas rejected on the basis of previous analysis?

We think the Mahoneys understate drastically the role that the experiment has played and will play in determining that negative income taxes can be administered and that they will not undermine work efforts. The impact of the experiment will take time because the basic structure of programs and the trade-offs are woefully misunderstood outside a small circle of analysts.

There is no question, of course, that the utility of the experimental data is limited. We have learned much more about this type of research since 1968, and future attempts certainly will resolve some of the problems in the New Jersey experiment.

But the experiment cannot be termed a failure. We think it has proved its worth in the area of program administration alone. The experience in designing administrative systems and the resulting data should aid both administrators of existing programs and those planning future programs. In fact, program administration itself would be a useful subject for future experimentation, and one that should be accorded a high priority.

Policy Implications: A Positive View

(By Michael C. Barth, Larry L. Orr, John L. Palmer)

The New Jersey–Pennsylvania experiment was designed and implemented with the primary objective of determining the labor supply response of able-bodied, prime-age male heads of families to negative income tax types of programs having various tax rates and guarantee levels.[1] For this reason, the labor supply results generally are regarded as the most critical policy aspect of the experiment. But the experiment also has considerable policy significance in two other areas. First, it

[1] The labor supply and most of the nonlabor supply findings of the experiment are relevant to more than a negative income tax program. Existing cash welfare programs such as AFDC UP and supplemental security income also have a structure involving guarantees and a tax imposed on other sources of income—as do such major in-kind welfare programs as food stamps and certain housing programs. Although we will use the term "negative income tax" in this paper, much of our discussion applies equally to any income-related transfer program having the same basic structure.

generated a substantial body of knowledge relating broadly to programs involving cash transfer whose level is scaled according to recipients' income. Second, it enlightened debate on the use and value of social experimentation as a tool of policy research.

We consider the policy impact and implications of the New Jersey experiment from a dual perspective. As economists, we are interested in better estimates of the parameters of labor supply. In addition, our positions provide us a vantage point from which to observe both the development of social experimentation and the utilization of the results of this particular experiment.

We also discuss the choice of a specific tax, or benefit reduction rate, in a negative income tax program—and the implication of the experiment for that decision. The criteria for evaluating social experiments also are developed and applied to the New Jersey experiment, along with an exploration of the implications for policy research emerging from this first, successful fielding of a controlled social experiment.

High-level administration officials have been involved in a detailed examination and an increasingly sophisticated discussion of the major issues involved in potential welfare reform policies and of evidence that could be brought to bear on them. The relation between work and welfare, particularly work disincentives, is among these issues. Many officials who originally believed that there would be large disincentive effects associated with high basic benefits or high benefit reduction rates were willing to revise their beliefs substantially in the face of the New Jersey experimental results and other relevant evidence. As a consequence, their willingness to give serious consideration to a program for which the working poor would be eligible has increased.[2] We see no reason to believe that, once the evidence has been more widely discussed, most congressmen and the general public would respond very differently.[3]

WELFARE POLICY: IMPLICATIONS AND IMPACT OF THE EXPERIMENT

Policy impact and policy implications, although closely related, are separable. Policy implications are abstract. They emerge from some portrait of reality that is imposed upon a particular policy issue most often one of program design. Policy impact concerns the actual effect of the experiment on the attitudes and behavior of policymakers and persons who influence them. Such effects are of most interest when they contribute to the promulgation or prevention of policy changes.

Labor Supply

POLICY IMPLICATIONS. We believe that the findings of the New Jersey experiment lend considerable support to the contention that, based upon existing evidence—that is, the results of the experiment viewed in

---

[2] One lengthy briefing on, and discussion of, the final results of the experiment included the secretary and under secretary of HEW, the under secretary of labor, and at least seven officials at the assistant secretary level from various executive agencies. Rarely does a research project receive this much interest from such high-ranking officials. Such exposure cannot help but improve the quality of subsequent discussion of related issues by these decisionmakers.

[3] There is already some evidence about Congress. Representative Martha W. Griffiths, in releasing Paper 13 of *Studies in Public Welfare* (prepared for the Subcommittee on Fiscal Policy of the Joint Economic Committee, "How Income Supplements Can Affect Work Behavior," stated: "A key obstacle to extending cash supplements to poor families headed by able-bodied men has been the fear that many will leave their jobs. The bulk of the evidence shows that such fears are unfounded. . . . The studies are in substantial agreement: a broad income supplement plan would add to the incomes of poor fathers, without causing such men to leave full-time work." JEC Press Release, February 18, 1974. The New Jersey experiment was among the studies cited by Griffiths.

conjunction with, and in relation to, other evidence on this issue—the best estimate of the effects of a national negative income tax type of program, one with tax rates and guarantees in the range tested, on the labor supply of prime-age able-bodied male heads of intact urban families are:

—First, in the aggregate, both in the short and the long run, the reduction in labor supply is likely to be quite modest, less than 10 percent, at worst.
—Second, any reduction, will be distributed across many workers rather than concentrated among a few.
—Third, the degree of reduction will not be very sensitive to the particular guarantee and tax rate chosen, at least among the lower tax rates.

These findings contain four important policy implications. First, public opposition to coverage of all intact families by an income-related cash-transfer program—to the extent that such opposition is based on fear of large reductions in work effort—should decrease. Second, the concern of policymakers about the disincentive effects of particular tax rates and guarantee levels should diminish. They can place heavier weight upon other criteria in the selection of an appropriate tax rate and guarantee level in any income-related cash transfer program.

Third, the case for a work test in an income-related cash transfer program covering intact families is weakened. In light of the administrative and other costs of a work test, the smaller the reduction in labor supply that would occur in its absence, the less cost effective it will be. In addition, whether a work test could prevent the small reductions that do occur is questionable. The fact that a work test for male heads of families is likely to be cost ineffective, however, does not necessarily make it undesirable. It may be necessary to preserve the integrity of the program.

Fourth, the very existence of the experiment as well as its results should raise the level of the policy debate surrounding work and welfare in general and the work-disincentive effects of income-related transfer programs in particular. Policy concern should be more explicitly articulated: for example, distinctions among responses of male heads, female heads, and secondary workers are more likely to be made. Debate now may center on acceptable amounts and kinds of labor supply response rather than on its presence or absence.

POLICY IMPACT. Because the final results of the experiment have been available for only a short time, we expect that most of the policy impact is yet to come. We have no way of knowing to what extent preliminary results reported to the House Ways and Means Committee may have influenced its chairman, Wilbur D. Mills, to support the family assistance plan. But it does seem likely that the negative findings at that time—that no abrupt or large reductions will occur in labor supply—might have allayed the concern of those who were not opponents of the program on other grounds.

Nonlabor Supply Policy Implications and Impact

In retrospect, the policy significance of the nonlabor supply aspects of the experiment appears to be at least as valuable as, if not more valuable than, that of the labor supply results. Many of the findings

are relevant not only to a new cash program for intact families, but also to any income-related program, including existing welfare programs. Most of these non-labor supply findings emerge from what Robert Levine terms the "demonstration aspects" of the experiment.[4] They pertain largely to administrative matters, but ones with important policy implications.

ADMINISTRATIVE FEASIBILITY. One apparent fear was whether a negative income tax program—or any other comprehensive federally administered income-related program—could ever be successfully administered. But more important than whether it can be done is how it should be done. The New Jersey experiment has helped to answer both questions.

The administrative lessons of the experiment are also relevant to other types of negative income taxes: for example, to a refundable tax credit that would replace personal exemptions. Administration of a refundable tax credit would involve detailed self-reporting, similar to that developed in the New Jersey experiment, by a population presently having minimal contact with the positive tax system.

THE ACCOUNTING PERIOD. Everyone can agree that an income-related transfer program should treat equally needy people equally. But even if one assumes that income is to be the measure of need, difficult issues are involved in the design of an equitable program. Are people with the same income in a given month but very different regularized annual incomes equally in need? Over what time period should equality of need be measured? Because there is no obvious answer to this latter question, the implications of alternative definitions of equal need must be examined.

The income maintenance experiment generated the first longitudinal data bases containing intrayear income flows for the low-income population, thus making possible the analysis of alternative periods of time over which income is to be counted to determine eligibility and benefit levels. Because the income of the low-income population generally fluctuates considerably within a year, both costs and coverage are highly sensitive to the length of the accounting period for a given guarantee. A given number of transfer dollars can be distributed in quite different ways depending on the length of the accounting period. Those with fluctuating, higher-than-average, but occasionally low monthly incomes are aided relatively more by a short accounting period; those with more stable monthly but lower average annual incomes are aided relatively more by a long accounting period.

In addition, there is continuing analysis and discussion of changing the accounting period in AFDC–UP and other income-related programs, most of which have a (nominal) one-month accounting period. Neither the acute awareness of this issue nor the data to permit its analysis would exist in the absence of the experiment.

INCOME REPORTING AND INFORMATION PROCESSING. The experiment generated considerable information on such issues as the trade-off in administrative cost between reporting at regular intervals as against reporting only when significant changes occur or between frequency of reporting and accuracy of the data reported. What is the ability of the population to report on a self-assessment basis, as in the U.S. income tax system? How much and what types of assistance are com-

[4] See Robert Levine's paper, "How and Why the Experiment Came About," in this volume.

patible with self-reporting? These are critical issues in the design of an administrative system for an income-related program. The reporting practices in present programs have not yet been much improved as a result, although evidence from the experiment has been brought to bear on the issue of increased frequency of reporting in the AFDC-UP program. But the experiment's results are having a major impact on the design of an administrative structure under current reform proposals.

Useful information with strong policy implications also was acquired about data processing needs of a regularized reporting and payment system, techniques of audit and verification, and the application of an assets test. As with income reporting, this information is having little impact on current program practices but is important in the design of reform proposals.

BEHAVIORAL RESULTS. Another major portion of the experimental analysis was devoted to exploring the effects of the negative income tax plans on economic, social, and psychological behavior and attitudes beyond the issue of the labor market.[5] With the exception of some aspects of consumption behavior, job search patterns, and educational attainment, the results of these studies were almost uniformly negative: no systematic pattern of significant experimental effect was found.

The policy implications of these results also are negative. If these findings are to be accepted at face value, marginal increments of income—increases in family income of about 25 percent in the experimental plans—will have no major impact on the life style and attitudes of low-income families. By the same token, such payments appear not to produce the deleterious effects sometimes associated with being on "the dole." If these basic propositions were to be accepted by all parties to the welfare reform debate, much extraneous rhetoric and emotional undercurrent could be dispensed with, and policy deliberation focused on more central issues.

Finally, we agree with Levine that, at least with respect to the New Jersey experiment, social experimentation is as much a political process as a detailed scientific one. We, too, think that a major result of the experiment probably will be "to make the negative income tax visible and therefore more feasible than it otherwise would have been."[6]

CHOICE OF A TAX RATE FOR POLICY

The results of measuring the differential effects of various guarantee levels and tax rates on labor supply fail to indicate that the different tax rates studies have significantly different disincentive effects.[7] It has been argued that such a finding supports the view that any income-conditioned transfer program for male-headed families should embody high tax rates in order to concentrate benefits among families with the lowest incomes.[8] Such a strategy, it is argued, will maximize

---

[5] The paper prepared by Peter Rossi, "A Critical Review of the Analysis of Nonlabor Force Responses," in this volume, contains a discussion of most of these analyses. These effects were not central to the purpose of the experiment and therefore received less attention than the labor supply effects. For this reason, the evidence mentioned below should not be considered definitive.

[6] Levine, "How and Why," p. 23.

[7] The experimental evidence for the 70 percent tax rate is relatively inconclusive because the benefits of the New Jersey and Pennsylvania AFDC-UP programs tended to dominate those plans with this high tax rate. Unless further analyses of the data are successful in disentangling these effects, the experimental results probably cannot be relied upon to predict the effect of tax rates in excess of 50 percent.

[8] See the paper by Bette S. Mahoney and W. Michael Mahoney, "Policy Implications: A Skeptical View," in this volume.

the antipoverty efficiency of cash transfers so long as the labor supply response to cash transfers is small. Although we agree that the experimental results have an important bearing on the choice of tax rate for public policy, we do not agree that the results necessarily argue for high tax rates.

The Goal of Transfer Policy

In the absence of any labor supply response, high tax rates certainly reduce the cost of bringing all low-income families up to a specified minimum income level. But it is not clear that the distributional objectives of transfer policy are simply to maximize the portion of total transfers going to the lowest decile of families or to minimize the cost of providing a specified minimum income. Ultimately, the goal of transfer policy is to change the shape of income distribution; thus, the optimal transfer policy is one that achieves the income distribution deemed optimal by policymakers. Viewed in this way, it makes no sense to speak of allocating a fixed transfer budget efficiently or setting transfer policy to minimize net transfers. The size of the transfer budget that policymakers ultimately will approve, large or small, will reflect the pattern of income distribution they prefer.

Distributional Effects of High and Low Tax Rates

What, then, are the distributional implications of the level of the tax rate?

High tax rates would have the effect of drastically compressing income differentials in the lower tail of the income distribution. In the extreme case, a program with 100 percent tax rates would virtually eliminate the gap between families with no private income and those just above the guarantee level.[9] At a minimum, this is certain to offend the sense of equity of those unaided—and quite possibly poor—families who end up near the bottom of the new income distribution; it may strike policymakers as inequitable as well.

In contrast, low tax rates would allow a much smoother compression of the income distribution over this range. The income differential between any two recipient families would be reduced by a fraction exactly equal to the tax rate—assuming no labor supply response. It seems quite plausible to us that policymakers might wish to preserve some degree of income differentials within the lower tail of the income distribution as a reward for private effort and initiative, even if work effort in itself is unaffected by the tax rate.

Basically, the choice of the tax rate is a question largely of distributional equity between the working poor and near poor whose incomes fall near the poverty line and the nonworking poor. At any reasonable budget level, a transfer program with a high guarantee and high tax rate would exclude from benefits large numbers of full-time, year-round workers whose earnings are low by any standard while providing relatively generous assistance to the families of unemployed or partially employed individuals.

In short, high tax rates are an efficient means of attaining a very specific type of income distribution, one in which all recipients have total incomes close to the guarantee level, and in which the remainder

[9] The distributional question on the positive tax side is analogous. A tax policy that imposed 100 percent marginal rates on incomes above a certain level would maximize tax collections from those best able to pay, just as 100 percent tax rates on the transfer side maximize payments to those most in need. Tax policymakers, however, have opted instead for a posttax income distribution that narrows rather than eliminates income differentials.

of the income distribution remains unchanged (except for changes caused by taxes required to finance the transfer). Whether this is an optimal distribution is a policy question that cannot be resolved by empirical evidence. What can be said on the basis of the experimental evidence is that, in making this distributional decision, policymakers need not be constrained by an expectation of induced labor supply response to tax rates in the range for which the experimental evidence is relevant.

Perhaps other adverse behavioral responses to high tax rates must be taken into account. For example, the incentives for divorce and desertion or for individual family members to leave the home may be a function of the tax rate. The abrupt change in marginal tax rates at the breakeven point also would create a variety of incentives for real or apparent shifts in income patterns over time and among sources of income. These incentives are similar to those involved in the infamous tax loopholes of the positive income tax; in the case of the negative income tax, families can largely escape taxation by clustering their income in periods when it is above the breakeven or switching to sources that are easily underreported.

Potential Effects of Extremes

But in viewing the tax-rate issue, the experimental findings do indicate some overall labor supply reduction in response to negative income tax payments—for secondary workers, a fairly substantial response. True, no clear and systematic variation in response appeared over the range either of tax rates tested or of guarantee levels employed, but the observed labor supply reduction must be in response to changes in one of these parameters, since they uniquely define the experimental treatment. Thus, we cannot say that either or both the tax rate and guarantee effects are zero, only that the experiment was unable to measure their separate effects reliably. Therefore, the potential for extreme values of these parameters to induce labor supply reduction cannot be dismissed out of hand, especially in light of the inconclusive evidence about the 70 percent rates.

As we have been reminded repeatedly, the tax rate that matters is not the rate applicable to a single program but the cumulative benefit reduction rate involved in the combination of program benefits that any individual receives. Even under a negative income tax that is fully integrated with the positive income tax system and replaces existing income-related cash transfers and food stamps, a variety of other income-conditioned tax and transfer programs are likely to remain: for example, the social security payroll tax, sales taxes, public housing, and premium contributions and benefits under national health insurance. If the cumulative benefit reduction or tax rate is to remain below 100 percent, the negative income tax rate must leave room for additional rates.

THE EXPERIMENT AS A TOOL FOR POLICY RESEARCH

By what criteria should anything as complex as the New Jersey experiment be judged? A nonexhaustive list would include the following:

—The central hypothesis should have been of compelling policy importance at the time the experiment was designed, with good reason to believe that this would continue to be the case.

—At the time of the experiment's design, there should have been no cheaper or simpler way of obtaining the desired information.[10]
—The experiment should have been competently and honestly managed—including, of course, the analysis.
—The tracks of the experiment should allow other social scientists, in principle, to replicate the operations and other analysts, in fact, to replicate the econometrics.
—The results, however complex, should speak directly to the initial hypothesis.

There is, we believe, little dispute that the experiment satisfies the first four criteria. A strong case for its meeting the fifth criterion also can be made, but there may be some disagreement. Having written the summary report of the experiment,[11] we are aware of the complexity of the results and of the paradoxes and loose ends. Complex results, however, often are produced by complex investigations of complex social phenomena. Even the initiators and designers of the New Jersey experiment did not expect it to provide definitive findings on the labor supply response of the population it covered. It was assumed in advance that no matter how apparently successful the execution of the experiment and how meticulous the analysis of the data, issues such as the potential Hawthorne effect and the methodological complexity, short duration, and small scale of the program would lead many people to question the relevance of its results for a national program. Nevertheless, it should be clear to everybody that the New Jersey experiment directly addresses the null hypothesis, that a negative income tax treatment would have no effect on labor supply.

What is troublesome about the experiment, particularly to those who must use its results, are precisely the paradoxes and loose ends. Because the results of the experiment are somewhat clouded at present by such factors as the AFDC–UP contamination and the inexplicably slow income growth of black families in the control group, the findings at this point are by no means beyond question. Thus, one cannot expect unanimity about the usability of the results in a policy context. Rather, one must rest upon a consensus of experts and consistency with other evidence. But is this really much different from the problems encountered in the utilization of any product of policy research? It is perhaps the costliness of this experiment and the presumed importance of its basic task, as much as the complexity of its results, that—quite rightly, in our view—lead to the calls for caution in generalizing from the results now available.

Usefulness of the Experiment

One evaluative criterion that is certain to be raised is the usefulness of the experiment. The fairness of this criterion deserves comment. Suppose Congress had enacted a universal cash transfer program in 1970. Would the labor supply results of the experiment have been useful? One can always argue that fine tuning of tax and guarantee parameters would always be important. Our guess, however, is that

[10] It has been suggested that analysis of data from ongoing federal social programs could have shed some light on those issues which the experiment was designed to examine. Although we have supported, do support, and will continue to support analyses of program data, we feel there were compelling reasons to seek better data. First, there is no control group for persons whose behavior generates the observation reflected in program data, and to a large extent those observations are self-selected. Second, the treatments of greatest interest are defined imperfectly at best and are uncontrolled in operating programs. Third, the cross-sectional data available from program records are grossly inferior to the longitudinal data obtained by special studies such as the experiment.

[11] U.S. Department of Health, Education, and Welfare, "Summary Report: New Jersey Graduated Work Incentive Experiment" (1973; processed).

the answer to the question of usefulness would then be negative. But surely this would not have been a fair criterion upon which to judge the experiment. Even though the labor supply results no longer would have been of compelling policy interest, the validation of the initial hypothesis alone suggests that the experiment performed its task.

Congress did not pass a cash transfer bill, but the administration currently is developing a comprehnsive proposal for welfare reform, and the results of the experiment already have been useful in policy discourse. The experiment's enduring contribution and its impact on any particular piece of legislation are for historians and journalists to argue.

Current policy developments aside, our model of the policy impact of the experiment runs as follows. In the short run the experiment may have little effect. If its results are not proved incorrect by competent authority, however, a strong presumption of a small disincentive effect probably will begin to prevail, in no small part because of the consistency between the results of the experiment and the cross-sectional studies.[12] Neither alone is sufficient to establish such a presumption. But this presumption is unlikely to produce a rapid change in the public stance of decisionmakers. A more probable outcome is that work-incentive effects will be given less weight in decisionmaking about transfer plans.

Longer-term attitudes may be affected in the following manner. Those who understand the methodology, assuming they are convinced of the accuracy of the experimental results, will begin to believe the results, which will represent a subtle shift in the conventional wisdom regarding the labor supply effects of transfer programs. Similarly, assuming professional approval, the results will begin to be accepted by the public.

Others may feel that this is too optimistic a view. But if policy research, which includes social experimentation, does not and cannot change the layman's views, why are we in this business? Our model of the utilization of the experimental results allows a very positive answer.

What the Experience Did Not Do

To this point, our evaluation of the experiment has considered only errors of commission. This is appropriate for two reasons: first, the criterion of compelling policy importance would not be satisfied if an experiment did not investigate a sufficiently important question; and, second, anything so complex as an experiment in the social sciences really can have only one hypothesis as its driving force.[13] Therefore, we feel that errors of omission are less consequential. Nevertheless, the experiment has been faulted for what it did not do, a criticism that deserves comment.

To expect that the experiment would answer all questions of policy interest regarding cash-transfer plans is to impose an impossible and unjustifiable burden upon it. Research, of whatever type, attacks targets of opportunity. The justification for a policy research project is that the research question be of policy importance and that the research can be done. On this criterion we give the experiment a good grade.

---

[12] See Irwin Garfinkel, "Income Transfer Programs and Work Effort: A Review," in *Studies in Public Welfare*, Paper 13.

[13] Felicity Skidmore's paper in this volume, "Operational Design of the Experiment," supports this point.

Moreover, the experiment does not and should not decide for policymakers whether to extend cash assistance to the working poor or at what levels and with what benefit-reduction rates. Indeed, no empirical evidence could do so. Research, no matter how relevant and competent, cannot tell us what national policy ought to be. It can provide some hard data as one input to the process that balances competing demands for scarce public resources.

Much more analysis of the experimental data certainly must be done. Only relatively simple models have been estimated to date, since the first task was to identify what happened. As time permits, a search for the "why," more explicit use of models of family labor supply, the joint determination of participation and hours, and other refinements will be required. Whether policy will move faster than analysis is a matter about which we can only guess. If the prospect of such an eventuality were to deter research, research would never get done.

IMPLICATIONS FOR POLICY RESEARCH

Among the longer-lasting effects of the New Jersey experiment may be those on governmental policy research. If one views policy research as an integral part of the policy process—and we do—then these effects may be regarded as policy implications or impact of the experiment. In the long run, we believe that these will prove to be substantial.

Whatever the technical defects of this particular experiment, it represents a marked departure from traditional techniques of research and evaluation, on the one hand, and, on the other, a monumental advance over demonstrations as a technique for testing behavioral hypotheses. As the first major controlled social experiment, it opened up an entirely new methodology of policy research. Like any other research tool, it must be used judiciously. But we believe that experimentation may be a superior approach whenever knowledge of individual behavioral responses is important to program design. Examples may be found throughout the whole range of consumption subsidies employed or contemplated by the government: in health, housing, social services, education, transport, and other areas. Of course, proposed experiments in any of these areas must be evaluated on their merits, including the degree to which they satisfy the five criteria set forth in the preceding section.

Prototype for Social Experiments

FEASIBILITY. We are only beginning to explore the potential of experimentation as a tool for policy research. It is far too early to assess the ultimate value of the experiments that have followed the New Jersey experiment, but almost certainly they would not have been possible without the pioneering example set in New Jersey. The change in the attitudes of governmental officials toward large-scale field research since 1967 has been striking. As Levine points out,[14] in 1967 HEW expected major political difficulties in the idea of the experiment, yet only 3 years later the department launched two new experiments in income maintenance—in Seattle-Denver and in Gary—on a scale that dwarfed the New Jersey project. And in 1973, only HEW's support saved a major experimental study of health insurance from being killed on political grounds by new leadership at the OEO.

[14] Levine, "How and Why."

There was nothing inexorable about the growing acceptance of experimentation as a means of policy research. Rather, it is attributable largely to the success of the experimenters in implementing an extremely difficult administrative undertaking in a manner that preserved fundamental analytical objectives. In the crudest terms, the experiment appeared to work without operation compromises that would have sacrificed its objectives. This reassured researchers that usable data could be obtained from field projects, and it reassured policymakers that such an undertaking need not be a political liability.

METHODOLOGY. The experiment not only demonstrated the feasibility of this new technique but also made major methodological contributions to its development. Perhaps the most basic and important was the notion that at least some social policies can be parameterized and a continuous response function estimated from observations generated in the field. This notion, so common in nonexperimental research, was virtually foreign to the literature or experimental design in the social sciences, which had been dominated by analysis of variance techniques. It was totally absent from the evaluation of methodology in previous demonstrations.

The idea of using regression analysis to estimate a continuous response surface as a function of well-defined policy parameters, combined with the recognition of the different costs of the various types of observations and the different policy interest in the various negative income tax plans, led to the development of the Conlisk-Watts sample allocation model. This statistical technique, which has formed the basis for sample design in most subsequent social experiments, offers great economy and flexibility in the design of experimental research—as compared, for example, with a balanced factorial design. It allows policymakers—or their policy research representatives—to focus their interest on specific areas of the policy domain without excluding less central policy options that may prove important in the end. It allows efficient allocation of resources over policy options and population groups with widely different budgetary costs; and it allows interpolation of results to points in the population domain not specifically included in the experimental design.

One can argue about the extent to which the potential of these statistical techniques has been realized in the New Jersey experiment. But it seems clear to us that they provide a focus of research objectives—in fact, an entirely new way of thinking about experimental design—that takes field tests of social policy out of the ill-defined realm of demonstrations and into the scientific realm of behavioral research.

DESIGN. In other important respects quite aside from the statistical contributions of the experiment, this project has formed a model for subsequent social experiments. The administrative and field procedures developed in New Jersey, including the carefully defined rules of operation, are a major contribution to experimental methodology.

We also have learned a great deal from the mistakes made in New Jersey. Among the more important lessons is that the design of social experiments must be robust with respect to exogenous forces and unforeseen events. The experimental treatments must be designed to insulate the participants both from existing programs that the policy under study is designed to replace and from unexpected legislative developments that threaten the integrity of the experimentally defined

policy parameters. In the second generation of experimentation, we are taking great care to build in this insulation.[15]

In addition, we now know that the design of the experiment must be robust with respect to analytical exigencies. The possibility of strong interactions between treatment variables and participant characteristics such as ethnicity dictates some tempering of rigid application of sample allocation models based on smoothly continuous response functions. Finally, although we hesitate to term the concentration of this experiment in a single geographical area a mistake, we have become acutely aware of the problem of generalizability of results. Much thought now is being given by government researchers to the problem of designing experiments that, although necessarily clustered in a small number of sites, yield results that can be confidently extrapolated to large geographical areas and altered market conditions.

Contribution to the Social Data Base

Apart from its importance as the prototype for social experiments, the New Jersey project has generated an exceedingly rich body of data that will be useful for investigation of a wide variety of behavioral issues wholly unrelated to the labor supply response to a negative income tax. It is virtually the only existing set of longitudinal data on the intrayear family-income dynamics of the working poor. Over $2 million of the $8 million cost of the experiment was devoted to the collection and preanalytical processing of the more than 15,000 interviews administered over the course of the experiment. To generate a comparable data base through surveys or collection of program data, even without the experimental variations, would have been equally expensive. Given the uniqueness of this data base and the care with which it was collected, we feel that at least this portion of the experimental expense is defensible on grounds wholly separate from the major purposes of the project.

More generally, the advent of social experimentation has helped to focus more attention on the inadequacies and defects of nonexperimental data. It was, after all, these defects which were a partial impetus to the New Jersey experiment. The biases attributable to self-selection in observations drawn from ongoing programs and private markets are now well recognized—in part because of the very attention they have received in the context of social experiments. Interest is growing in the use of longitudinal data to study dynamic processes, and with it an increased awareness of the dearth of reliable longitudinal surveys. To parse out the contribution of experimentation to these developments is, of course, impossible. Our own judgment is that it has been consequential.

Contribution to Research Talent

Finally, social experimentation in general and the New Jersey experiment in particular have helped to create a community of academic researchers with an abiding interest in the policy issues of income maintenance. It has provided a focus—and, to some extent, funding—for their

---

[15] In the Seattle and Denver experiment into income maintenance, all treatments are designed to dominate existing welfare benefits. In the health insurance experiment, a rather elaborate set of rules and compensatory cash payments has been developed to preclude receipt of benefits from nonexperimental health insurance plans, both public and private, although still guaranteeing the participants' eligibility for such benefits at the end of the experiment. The use of compensatory cash payments has quite general applicability for insulating experimental treatments involving in-kind benefits and earmarked cash transfers.

research efforts, as well as data for their models. In the process of working through the development of a model negative income tax program, theoretical and programmatic issues that otherwise might have remained hidden have been brought into the light for analysis. Conversely, that same process has helped to broaden the perspective of the researchers themselves, to include programmatic consideration that do not emerge naturally from theoretical models. Those of us who mediate between the research community and the policy process are in a position to appreciate both the creation of this reservoir of talent and that active interest in income maintenance policy which its members have maintained.

SUMMARY

The New Jersey experiment has numerous policy implications and already has had significant policy impact. Assuming that the analysis to date is not contradicted, the experiment suggests that the work-disincentive response of prime-age male family heads in the urban Northeast to an income-related cash transfer would be quite modest. In our opinion, this result, together with the results of cross-sectional studies of labor supply behavior, will begin to establish a presumption of a small labor supply response of male family heads under a national program of similar characteristics. If we are correct, then a rather remarkable shift in informed opinion will have taken place since the 1966–68 period during which the experiment was proposed, designed, and begun.

One of the more important findings of this experiment is that no different labor supply response across tax rates was detected. The policy implication of this finding represents a loosening of constraints on program designers: that in making decisions on the desired income distribution, policymakers can select benefit reduction or tax rates to satisfy a broad range of social goals, not only the goal of minimizing reduction of work effort.

Assessing the policy impact of the experiment at this point is at least as hazardous as deriving its policy implications. Still, there seems to be general agreement that the experiment had some impact on the deliberations about the family assistance plan, at least in the House of Representatives. More recently, the final report of the experiment coincided with a substantial effort at the staff and policy levels to reexamine a range of welfare-related issues. Perhaps a cautious policy analyst would have waited a few years for a reanalysis of the data before presenting the results to policy-level officials. For good or ill, the results of the experiment, together with other relevant information, have been heard and, we believe, have had some impact on policymakers. Precisely how much and whether the experimental results will affect the existence and nature of a new welfare reform proposal are matters for speculation.

The uniqueness of this particular experiment should be noted. It was, of course, the first, and our positive assessment of the policy value of this experiment results in part from this fact. For example, the experiments that quickly followed—the rural, Seattle-Denver, and Gary income maintenance projects and others on housing allowances, performance contracting in education, and health insurance—no matter how successful they may be, may not have the same impact on future policy research.

What seems most important at this time is that a new methodology of policy research was developed successfully and directed to an important question. Answers, however complex, were produced and heard by persons in a position to act on the information. It is very early in the half-life of the knowledge generated by the experiment to say much more.

III. Excerpts From: Welfare in Rural Areas: The North Carolina-Iowa Income Maintenance Experiment*

AN OVERVIEW

(By Larry L. Orr)

The rural income maintenance experiment is the second of four major experiments to test the consequences of a universal income-conditioned cash transfer program. Its uniqueness lies in its focus on the rural sector. The New Jersey experiment yielded a great deal of information about the effect of various negative income tax plans on behavioral and attitudinal characteristics of urban wage earners.[1] But these results were not expected to be directly applicable to the rural sector, in which over one-third of the nation's poor reside. Differences in the work responses of rural and urban residents to such a program were expected because of differences in alternative employment opportunities and in the proportion of self-employed people. An accurate estimate of the size of any work disincentive, both rural and urban, was necessary to estimate the cost of a nationwide program.

Also, it seemed likely that there would be some features in a program designed for addressing urban poverty problems that were not suited for rural poverty. For example, a large number of rural residents with low incomes are operators of farms or businesses in small towns. Determination of annual income as well as the appropriate timing of payments are different for the self-employed than for wage earners. This is especially true for those farmers who receive their entire annual income at harvest time. The provisions for self-employed individuals in the New Jersey experiment were admittedly simple and probably inadequate for a nationwide comprehensive negative income tax program.

The New Jersey experiment restricted eligibility to families of two or more members, with an able-bodied male head between the ages of eighteen and fifty-eight. Since a large number of poor households are headed by females of working age, a study of their work behavior was considered desirable by the Office of Economic Opportunity (which funded the project). The second major category excluded from the New Jersey experiment was composed of those of retirement age. Men and women sixty-five years of age and over made up 16 percent of the

[1] The results of this experiment were reviewed and evaluated in Joseph A. Pechman and P. Michael Timpane, eds., *Work Incentives and Income Guarantees: The New Jersey Negative Income Tax Experiment* (Brookings Institution, 1975). Results from the remaining two experiments in Gary, Indiana, and Seattle-Denver are not yet available.

*Published 1978 by the Brookings Institution.

poor people, and headed about one-third of the poor households, in the United States. While the work incentive issue was less significant for this group, it was nevertheless considered important.[2]

The need for experimentation in a rural setting in conjunction with the urban experiment in New Jersey led to a planning grant from the Ford Foundation to the Institute for Research on Poverty at the University of Wisconsin. Under the grant Poverty Institute staff members representing the disciplines of economics, agricultural economics, sociology, political science, law, and social work joined in an interdisciplinary effort to design the rural experiment. The experiment was patterned after the one in New Jersey: it had the same basic objectives, a similar experimental design, and identical duration. It differed from the urban experiment in that it extended eligibility to single households as well as to those headed by females and the aged. Minor variations also existed in the definition of earned income and in the accounting period for determining income and payments due.

Subsequent papers in this volume provide a detailed review and critique of the design, operation, and findings of the experiment. In this paper the purpose is to present a broad overview.

DESIGN AND METHODOLOGY

The rural experiment was conducted over the three-year period 1970–72 in rural counties of Iowa and North Carolina. The experimental sites were purposively selected to be representative of five states in the South and three in the Midwest that contain about one-third of the U.S. rural poverty population. All participating families in the Iowa sample were white; in North Carolina the sample was about evenly divided between black and white families.

The experimental treatments included five different negative income tax plans, each characterized by a basic benefit, or guarantee, and a benefit reduction, or tax, rate.[3] Guarantees ranged from 50 to 100 percent of the poverty line, and tax rates ranged from 30 to 70 percent. In addition, a control group that received no payments, but was interviewed periodically along with the treatment groups, was also enrolled.

A total of 809 families was initially enrolled; of that number, 729 remained enrolled for the entire three years of the experiment. Families were selected randomly within the experimental sites and eligible families were randomly assigned to treatment plans and the control group. To be eligible for selection, families were required to have incomes in the year prior to the experiment that were less than 150 percent of the

---

[2] The female-headed portion of the sample, numbering 108 (less than fifty-eight years of age) added breadth to such studies as those of nutrition, health, and children's school performance, but by itself was too small to provide a comprehensive answer to the question of the negative income tax effect on the labor supply of female heads. The same may be said of the older heads (114 at enrollment) and the impact of the negative income tax on retirement decisions. Policymakers must await the results of other experiments to gain meaningful insight into these issues.

[3] The guarantee is the benefit that would be received by a family with no other income; the tax rate is the rate at which benefits are reduced as other income rises.

official poverty line. Selections were stratified by family type to obtain subsamples of 587 intact nonaged husband-wife families, 108 families with female heads, and 114 families with an aged head. The results reported here are primarily for the husband-wife subsample. The initial allocation of the 269 families in the experimental group among the five treatment plans is shown in the following tabulation:

	Tax rate (percent)		
	30	50	70
Guarantee (percent of poverty line):			
50		37	
75	67	75	30
100		60	

There were 318 families in the control group.

TABLE 1.—AVERAGE ANNUAL INCOME AND PAYMENT, NONAGED HUSBAND-WIFE FAMILIES, BY SITE, OVER COURSE OF THE RURAL INCOME MAINTENANCE EXPERIMENT, 1970–72

[In dollars]

Type of income	Wage earners			Farmers	
	North Carolina				
	Black	White	Iowa	North Carolina	Iowa
Total income [1]	5,692	5,544	7,364	5,649	5,676
Wage income [1]	5,460	5,280	6,568	3,260	931
Net farm income [1]				2,615	4,882
Negative income tax payments	1,574	1,560	1,343	1,723	1,534

[1] Control group mean.

Sources: Wage earner income data are from U.S. Department of Health, Education, and Welfare, "The Rural Income Maintenance Experiment: Summary Report" (HEW, November 1976; processed), p. 18; wage earner negative income tax payments are from the files of the Institute for Research on Poverty; other data are from Wendell E. Primus, "Farm Work Response of Farm Operators," in D. Lee Bawden and William S. Harrar, eds., *Rural Income Maintenance Experiment: Final Report* (University of Wisconsin-Madison, Institute for Research on Poverty, 1976), vol. 4, ch. 1, pp. 63, 66.

Payments were made biweekly on the basis of monthly reports of income and family size filed by the families. For the self-employed, income was defined net of business expenses. Average family income and experimental payments over the course of the experiment for various subgroups of the nonaged husband-wife sample are shown in table 1.

In addition to the income reports, interviews were conducted quarterly with household members fifteen years of age and over to gather attitudinal and behavioral data. Information was also gathered from schools and other public organizations.

The effects of the experimental treatments were estimated through regression analysis.[4] A number of different outcome variables (as measured by the quarterly interviews) were analyzed as dependent variables. In each case, a set of control variables, such as age, education, and family size, was entered to standardize for any nonexperimental differences among the treatment groups. In most cases, the preexperimental value of the dependent variable was also included as a control variable, both to account for any preexperimental differences among treatment groups in the level of response and to improve the precision of the estimates of experimental effects.

The experimental treatments themselves were characterized with a set of three independent variables denoting whether the family was in a treatment or control group and the levels of the guarantee and tax rate of the plan to which it was assigned. In addition, in some analyses the treatment variables were interacted with various measures of family and individual characteristics to test whether the response to the experiment varied systematically across families.

The measures of experimental response presented in this overview are, in most cases, differences between treatment and control families obtained by evaluating the response function for treatment families for a plan with a 45 percent tax rate and a guarantee at 80 percent of the poverty level of income.[5] These values were chosen to represent the weighted average of the five experimental treatments.

VALIDITY AND INTERPRETATION OF FINDINGS

Several factors bearing on the validity, interpretation, and generalizability of the findings of the rural experiment should be noted.

When the experiment began in 1970, about 35 percent of the U.S. poverty population lived in rural areas (on farms and in towns of 2,500

[4] The estimation technique was an error components pooling method. Quarterly observations were pooled for wage earners, and annual observations pooled for farmers.

[5] In some of the analyses where tax and guarantee effects were insignificant, the reported differentials are based on a simple treatment/control dummy variable.

or less). It was administratively infeasible, however, to draw a sample that represented the entire U.S. low-income rural population. Instead, two sites were selected that appeared to be typical of two major areas of concentration of the rural poor—the South and the Midwest. While the samples in these sites are not a strict statistical probability sample of even those areas, they are probably reasonably representative of the five southern states (Alabama, Georgia, Mississippi, North Carolina, and South Carolina) and three midwestern states (Illinois, Iowa, and Wisconsin) from which they were selected. These states included about one-third of the U.S. rural poverty population in 1970. The results of the experiment cannot be generalized with the same confidence to other rural areas, or to ethnic groups not included in the experimental sample.

A second significant feature of the experiment was that, like that in New Jersey, it lasted only three years. The response of participants in a permanent program might be somewhat different. There are theoretical reasons for believing that the observed work response to the guarantee may be understated and the observed response to the tax rate may be overstated compared to that in a permanent program. While these biases tend to be offsetting, they are of concern. More information on the extent of the bias, if any, should come from the Seattle-Denver experiment, where variation in the duration of the experiment was explicitly introduced as an experimental variable.

Third, no work requirement was imposed; participants did not have to register for work or accept offered employment to receive payments. Observed reductions in work and income may therefore be greater than those under an income maintenance program with a work requirement.

Fourth, relatively few families were assigned to plans at the 50 percent guarantee and at the 70 percent tax rate. Consequently, generalizations about the effects of low guarantees, or about high tax rates, should be made with considerable caution.

Fifth, sample attrition was remarkably low for a three-year panel study: only 9.9 percent of the families dropped out (voluntarily or involuntarily) during the entire period. According to an analysis of those who dropped out, attrition should not have led to any appreciable bias in estimates of work response to the experiment.*

---

* See Glen G. Cain and Steven G. Garber, "Attrition," in D. Lee Bawden and William S. Harrar, eds., *Rural Income Maintenance Experiment: Final Report* (University of Wisconsin—Madison, Institute for Research on Poverty, 1976), vol. 2, chap. 1 (hereafter *Rural Income Maintenance Experiment*).

Finally, it should be noted that neither Iowa nor North Carolina has an aid to families with dependent children—unemployed parents (AFDC-UP) program, so that there was no confounding of treatment effects due to a cash welfare program for the same population, as there was in the earlier New Jersey experiment.[7] Sample families were eligible to receive unemployment compensation benefits and food stamps but participation in these programs was quite low.

BEHAVIORAL RESPONSES TO THE EXPERIMENT

In this section, the major behavioral responses, as estimated by the rural experiment analysts, are presented. Income and work responses were estimated separately for wage earners and farmers. In addition, a number of other types of response were analyzed, including several forms of expenditure and consumption, job change and job search, geographic mobility, psychological well-being, marital stability, political participation, and the attitudes, behavior, and school performance of children.

Income and Work Responses of Wage Earners

Separate analyses of income and work responses were performed for families of wage earners and those of farm operators in North Carolina and Iowa. The wage earner sample was limited to husband-wife families of constant marital status, where the husband was less than sixty-three years old and not disabled, and where the primary source of income was not from self-employment activities. A total of 264 families met these criteria. 146 in the control group and 118 in the treatment groups. The allocation of the 118 treatment families among plans is shown below:[8]

	Tax rate (percent)		
	30	50	70
Guarantee (percent of poverty line):			
50		5	
75	36	40	7
100		30	

Approximately one-half of the wage earner sample was composed of North Carolina blacks, about one-fourth were North Carolina whites, and one-fourth Iowa whites.

Within the wage earner sample, separate analyses were performed for husbands, wives, and dependents as well as for the family as a whole.[9] Measures of income and work responses that were analyzed as dependent variables included total income (excluding public assistance and experimental transfers), wage income, hours worked for wages, and employment status (or number of earners).

[7] For a discussion see Henry J. Aaron, "Cautionary Notes on the Experiment," in Pechman and Timpane, eds., *Work Incentives and Income Guarantees*, pp. 88–110.

[8] The analysis was performed on pooled quarterly observations so that there were approximately twelve times the number of observations as the number of families shown.

[9] See the four papers by D. Lee Bawden in *Rural Income Maintenance Experiment:* "The Analytical Approach to Measuring Work and Income Response of Wage Earners"; "Income and Work Response of the Family"; "Income and Work Response of Husbands"; and "Income and Work Response of Wives and Dependents," vol. 3, pt. 1, chaps. 1–4, respectively.

TABLE 2.—EFFECT OF RURAL INCOME MAINTENANCE EXPERIMENT ON INCOME AND WORK RESPONSES OF NONAGED HUSBAND-WIFE WAGE EARNER FAMILIES, BY SITE [1]

[Percent of control mean]

Wage earner and variable	North Carolina		Iowa [2]	8-state aggregate [3]
	Black [2]	White [2]		
Husband:				
Wage income	[a]−7	[a]0	[a]−10	−4
Wage hours	−8	[b]+3	−1	−1
Employment rate	−1	−1	0	−1
Wife:				
Wage income	[b]−41	−3	−32	−25
Wage hours	[b]−31	−23	−22	−27
Employment rate	[c]−25	[a]−28	[a]−38	−28
Dependent:				
Wage income	−19	[a]−57	−8	−39
Wage hours	−16	[a]−66	−27	−46
Family as a whole:				
Total income	[b]−14	[b]−9	[b]−18	−13
Wage income	[b]−14	[a]−8	[b]−17	−12
Wage hours	[a]−10	[a]−18	−5	−13
Number of earners	[b]−6	[b]−16	[a]−8	−11

[1] Significance levels, based on joint F-test on coefficients of treatment dummy and tax and guarantee variables, are indicated as follows:
[a] Significant at the 0.05 level.
[b] Significant at the 0.01 level.
[c] Significant at the 0.10 level.
[2] Responses are standardized to a 45-percent tax and 80-percent guarantee plan.
[3] Weighted averages of the basic data from which the subsample percentages were derived, using the following weights: North Carolina blacks, 0.31788; North Carolina whites, 0.48943; Iowa, 0.19269. No tests of significance can be computed for these differentials. The 8 States from which the experimental sites were selected are Alabama, Georgia, Illinois, Iowa, Mississippi, North Carolina, South Carolina, and Wisconsin.

Sources: U.S. Department of Health, Education, and Welfare, "The Rural Income Maintenance Experiment: Summary Report" (HEW, November 1976; processed), p. 38, and computer printout in the files of the Institute for Research on Poverty.

The experimental effects on these measures, as estimated by regression analysis, are shown in table 2.[10] The first three columns of this table show responses by race/site subgroup. The fourth column gives the combined weighted-average response, with weights chosen to reflect the racial and regional composition of the rural, nonfarm, low-income population in the eight states from which the experimental sites were selected.

[10] The tests of significance shown in table 2 are joint F-tests on the three treatment parameters in the regression, not tests of the significance of the treatment/control differential at the point at which the response surface is being evaluated. For a more direct test of the significance of the treatment/control differential, see Department of Health, Education, and Welfare, "The Rural Income Maintenance Experiment: Summary Report" (November 1976; processed), tables 3–6, where a simple treatment dummy specification is employed.

The experimental responses of wage earners varied substantially among family members, and, to a lesser extent, among measures of response and among sample subgroups. In general, the responses of husbands were much less significant than those of wives and dependents. However, the relatively large percentage reductions in income and work of wives and dependents were measured against a small base. The wage income of wives accounted for only 5–21 percent of total family wage income in the three subgroups, while wage income of dependents was less than 10 percent of total family wages. Husbands showed reductions in wages or hours in only two of the three subgroups, and in none of the three groups did any noticeable withdrawal from employment occur.

When the income and work responses of all family members are combined, a somewhat more consistent pattern emerges. In two of the three groups, wages and total family income (excluding experimental payments) fell by 14 to 18 percent, while hours and employment were reduced by 5 to 10 percent. Among the third group, the pattern is reversed, with a reduction in hours and employment of 16 to 18 percent and a reduction in wages and total income of 8 to 9 percent.

The combined weighted responses for the three groups show a much more uniform response across the various measures. As shown in the last column of table 2, the weighted income and work reductions of husbands ranged from 1 to 4 percent; for wives, 25 to 28 percent; and for dependents, 39 to 46 percent. For the family as a whole, all weighted response measures fell by 11 to 13 percent.

While the detailed responses of the three subgroups give a useful indication of the variability and statistical significance of the various responses, the weighted responses give a better summary of the overall pattern of response. On the basis of these estimates, we can conclude that for rural wage earners a negative income tax of the type and level considered here will have little or no impact on the employment rate and earnings of husbands, but that it will cause about a one-fourth reduction in the employment rate of wives and a decline of nearly one-half in the employment rate of dependents, with concomitant reductions in their hours and earnings. Overall family income—which determines the level of benefit payments and net cost to the government—would fall by about 13 percent.

Analysis of the variation of response relative to the levels of the guarantee and tax rate yielded mixed results. In general, the level of the guarantee had no effect on the size of the response. About half of the response measures for the three subgroups, however, did appear to be sensitive to the tax rate in the expected direction. The most statistically significant tax rate effects were found for family and husband income measures for the two North Carolina groups, where an increase in the tax rate of 10 percentage points resulted in increases in the treatment/control differences of 8 to 21 percent of the control mean.

Various analyses of interaction variables were performed to test whether the experimental response varied with family or individual characteristics. Among black husbands in North Carolina, for example, the response declined significantly with age. In general, none of these differences in response proved to be statistically significant although there was some evidence of a greater response among North Carolina husbands working as hired farm workers, wives either with school-age

children or whose families engaged in some farm work, and dependents eighteen to twenty years of age. The response of wives also varied seasonally, with the largest response during the winter months when employment rates were lowest in both treatment and control groups.

Income and Work Responses of Farmers

In the first year of the experiment, 262 families reported some hours of work devoted to operating or managing a farm. After exclusions for such factors as negligible or discontinuous farming activities, extreme age, and changed marital status, about 220 families remained in the sample.[11] The allocation of the 117 farm operators of the experimental group among treatment plans is shown below:[12]

	Tax rate (percent)		
	30	50	70
Guarantee (percent of poverty line):			
50		24	
75	17	28	14
100		24	

The farm operators were about evenly divided between the two sites. The principal measure of income analyzed for farm families was farm profit, defined as gross revenue less current operating expenses.[13] This definition of profit includes gross returns to fixed factors of production (land and capital), as well as to the operator's own labor. Average farm profits in the control group were substantially higher in Iowa ($11,895 a year) than in North Carolina ($4,758 a year). In both sites, the experimental treatments appeared to reduce farm profits—by 25 percent in North Carolina and by 8 percent in Iowa—although these reductions had low statistical significance (20 percent and 15 percent, respectively). Changes in the tax rate or income guarantee level had no significant experimental effect, and there was no distinct time trend over the three years of the experiment.

The observed reduction in farm profits may be partly the result of underreporting of farm income in the experimental group. Farm income was seriously underreported on the income report forms used to calculate payments and while the interview data showed considerably higher farm income, it too may reflect some systematic underreporting.

Several measures of labor supply were analyzed for farm families. Since 78 percent of farm families in North Carolina, and 50 percent of those in Iowa, had one or more members who worked for wages, effects on both farm work and wage work were estimated. The labor supply results for farm operators and their wives are shown in table

[11] The exact sample size varied among analyses, as different criteria were applied to define the appropriate sample.

[12] Lynne Fender, William S. Harrar, and Brian Kastman, "Sample Selection and Description," in *Rural Income Maintenance Experiment*, vol. 1, chap. 4, table 3. There were 110 families in the control group.

[13] See Lewis T. Evans, "Relative Economic Efficiency of Farms," in *Rural Income Maintenance Experiment*, vol. 4, chap. 4.

3.[14] In both sites, operators and their wives considerably reduced their hours of wage work under the plan although in Iowa both spouses showed an increased probability of employment in wage work.[15] Hours of farm work by farm operators, on the other hand, increased by about 11 percent in both sites; however, only the increase in North Carolina is statistically significant. The net result—in total hours of work, for both farm operators alone and for operators and wives—was a decline in North Carolina, and an increase in Iowa. These results, particularly those for wage work, must be viewed with some caution because of the small number of operators and wives who actually worked for wages, and the generally low statistical significance of the estimates.

TABLE 3.—EFFECT OF RURAL INCOME MAINTENANCE EXPERIMENT ON FARM OPERATORS' AND WIVES' LABOR SUPPLY, BY SITE

[Percent of control mean]

Worker and labor supply measure	North Carolina	Iowa
Farm operators:		
Hours of farm work	[1] 10.7	10.9
Hours of wage work	[2]—31.3	[2]—10.0
Employment in wage work	—6.0	25.6
Total hours of work	—2.7	9.5
Wives:		
Hours of wage work	[2]—62.7	[2]—53.5
Employment in wage work	—8.2	7.0
Farm operators and wives: Total hours of work	[2]—16.4	[2] 7.3

[1] Significant at the 0.10 level.
[2] This differential is derived from other estimates; therefore no significance levels can be computed.

Source: Stuart H. Kerachsky, "On Farm-Off Farm Work Decisions," in *Rural Income Maintenance Experiment*, vol. 4, ch. 2, table 12.

It may seem surprising to find a positive experimental effect on hours of farm work, but the effect of a negative income tax on the labor supply of farm operators is not theoretically unambiguous as it is for wage earners. Farm operators have the opportunity to change the mix of land and capital they employ as well as the amount of their own labor; they can easily shift from wage work to farm work. This might account for part of the observed increase in farm hours except that the increase was not significantly larger for farmers who worked for wages than for those who did not.

It is also possible, of course, that the observed increase in farm hours merely reflects systematic overreporting of hours by farmers in the experimental group. Hours devoted to farm work are not as easily defined or measured as wage hours, and may therefore be subject to serious re-

[14] See Stuart H. Kerachsky, "On Farm-Off Farm Work Decisions," in ibid., chap. 2.
[15] It should be borne in mind that these estimates are based on annual data, so that the employment variable measures the probability of working for wages at any time during the year, rather than at a given point in time.

porting errors. However, the only obvious reason for experimental families consistently to report more farm hours than control families is that the reduction in their wage work resulted in more nonmarket time that could be attributed to farming activities. Reporting bias is therefore difficult to distinguish from a real shift from wage work to farm work; moreover, overreporting should not account for the increase in farms hours of those who did not work for wages.

Furthermore, direct analysis of production activities showed no substantial shifts in the composition of farm output toward labor-intensive goods.[16] The same analysis did indicate an experimentally induced drop in production, consistent with the finding of lower profits.

The experimental effect on farm hours did not differ significantly with variations in the levels of the tax rate or the guarantee. There was, however, a marked time trend, especially in North Carolina, with the size of the response growing over the course of the experiment. Analyses of interaction variables indicated that the increase in farm hours was largest for younger operators with smaller families and a smaller proportion of rented land.

The simultaneous findings of decreased profits and output, and increased hours of farm work among operators, imply that efficiency of farm operations declined among treatment families and direct analysis confirms this fact.[17] Both price efficiency—use of the optimal combinations of inputs—and technical efficiency—the amount of output produced from a given combination of inputs—were analyzed. Farmers in the treatment groups were found to be less technically efficient than those in the control group, while price efficiency appeared to be unaffected. The differences in technical efficiency were most pronounced in North Carolina, where the differential increased over the three years of the experiment. Furthermore, the decrease in technical efficiency was significantly greater at higher tax rates.

Effects on Consumption, Assets, and Debt

During the periodic interviews a variety of data were collected on patterns of expenditure and holdings of assets and debt. Analyses were performed on the experimental effect on nutrition, homeownership and housing expenditures, clothing purchases, use of medical care, ownership of consumer durables and liquid assets, and levels of loan and store debt. The major findings of these studies are briefly described below.

NUTRITION.[18] Dietary intake data for a 24-hour period were collected from 612 families in the third quarterly interview, and from 710 families in the eleventh quarterly interview. From these data, indexes of consumption of ten basic nutrients, expressed as a percentage of the recommended daily allowance, were formed. A combined index, the mean adequacy ratio (MAR), was defined as the arithmetic average of the ten individual indexes. In North Carolina, where control families scored 79 on the MAR scale, significant positive treatment effects of 3.0

[16] See William E. Saupe, "Farm Business Decisions," and Wendell E. Primus, "Farm Work Response of Farm Operators," both in *Rural Income Maintenance Experiment*, vol. 4, chaps. 5 and 1, respectively.

[17] See Evans, "Relative Economic Efficiency of Farms."

[18] See J. Frank O'Connor, J. Patrick Madden, and Allen M. Prindle, "Nutrition," in *Rural Income Maintenance Experiment*, vol. 5, chap. 6.

and 3.6 percentage points were found in the two quarters, as well as positive effects for nine of the ten individual nutrients. [19] In Iowa where nutritional adequacy was higher initially (a MAR of 89 for the control group), no significant effects on nutrition were found.

HOUSING.[20] Among the 321 families who had not purchased a home prior to the experiment, the probability of buying a home during the three years of the experiment was about 0.06 higher among families in the treatment group than the control group, with most of the differential effect attributable to North Carolina families. Among the 55 families who bought a home during the experiment, treatment families appeared to make the purchase two or three years earlier in their life cycle, with the differential being significant for farmers but not, in general, for other families. No significant difference in purchase price was detected, nor was there any significant experimental effect on rents paid by families who moved during the experiment.

CLOTHING.[21] Clothing expenditures in the winter months of each year were analyzed separately for two-parent families, both those in which the husband had wage income in every year and those in which the husband did not. Families of wage earners spent $0.025 out of every additional dollar of experimental payments on clothing, as compared with $0.05 from an added dollar of wives' income and $0.007 from other income earned by the family (including the husband's). Total clothing expenditures of nonwage earners' families were not significantly affected by the experimental payments.

MEDICAL CARE.[22] No significant experimental effects were found on the use of medical care, as measured by expenditures, medical visits, and possession of health insurance, or on the state of health, as measured by bed-days, work-loss days, and presence and severity of chronic conditions.

CONSUMER DURABLES AND CARS, LIQUID ASSETS, AND DEBT.[23] Increases in family income in the form of income maintenance payments may cause the family to alter the level of its holdings of assets and debts. Because these adjustments may require several years to complete, estimates were made of both the short-run—that is, current year—and complete long-run adjustment of assets to transfer payments. Separate analyses were performed for black and white families in which the husband had wage earnings, and for farmers. The estimated long-run effects on holdings of various types of assets as well as debts for wage earners and farmers are shown in table 4. The estimation technique does not allow confident calculation of statistical significance for these figures. In general, however, it appears that the effects for black wage earners were more statistically significant than those for white; the estimates for the farm sample were generally not significant.

---

[19] A score of 67 on the MAR scale is deemed nutritionally inadequate or dangerous to health. The individual scales were truncated at 100.

[20] See Aaron C. Johnson, Jr., "Housing Consumption," in *Rural Income Maintenance Experiment*, vol. 5, chap. 1.

[21] See Christine J. Hager and W. Keith Bryant, "Clothing Expenditures," in ibid., chap. 4.

[22] See Stuart H. Kerachsky, "State of Health and the Utilization of Medical Care," in ibid., chap. 5.

[23] See W. Keith Bryant and Christine J. Hager, "Consumer Durables, Cars, Liquid Assets, and Debts of Wage-Working Families," in ibid., chap. 2; and Bryant, "Consumer Durables, Cars, Liquid Assets, Short-term Farm Capital and Nonreal Estate Debts of Farm Families," in ibid., chap. 3.

TABLE 4.—LONG-RUN RESPONSE OF ASSETS AND DEBTS TO A NEGATIVE INCOME TAX, BY FAMILIES OF WAGE EARNERS AND FARMERS

[In dollars]

Type of family and race	Consumer durables	Cars	Store debt	Loan debt	Liquid assets
Wage earner:					
Black	168	167	−12	658	42
White	170	130	57	−1,245	187
Farmer	122	−87	−60	−268	19

Sources: W. Keith Bryant and Christine J. Hager, "Consumer Durables, Cars, Liquid Assets, and Debts of Wage-Working Families," in *Rural Income Maintenance Experiment*, vol. 5, ch. 2, tables 411.1, and 411.5 through 411.8; and Bryant, "Consumer Durables, Cars, Liquid Assets, Short-term Farm Capital and Nonreal Estate Debts of Farm Families," in ibid., ch. 3, p. 22.

The total experimental effects for wage earners shown in table 4 are based on regression equations that included a binary treatment variable, denoting eligibility for experimental payments, and a measure of the amount of payments a family would receive at its "normal" income level. In some cases these two effects were offsetting. For example, for black and white wage earners, simple eligibility for payments appeared to reduce loan debt by $2,638 and $1,540, respectively, while increases in the size of payments increased loan debt. Unfortunately, the estimation technique did not allow the analysts to distinguish between the effect of eligibility and initial differences in stocks held by the treatment and control groups. Thus, in those cases where the overall effect is dominated by a large eligibility effect—particularly the large decline in loan debt for white wage earners—considerable caution should be used as they may partly reflect initial differences in holdings rather than treatment effects.

Other Effects

Analyses were also performed on the experimental effects on job change and search, geographic mobility, psychological well-being, family structure, political participation, and various aspects of behavior of youth. Their results are briefly summarized below.

Job Change and Job Search [24].—Job turnover, duration of unemployment, and job selection were analyzed for male wage earners who were employed at the beginning of the experiment. The experimental payments appeared to have no overall effect on the probability of leaving a job, though significant effects were found for some subgroups of workers. For instance, those in experimental groups who initially had more desirable positions were less likely to leave their employers than similar people in control groups, while those in experimental groups with less desirable positions were more likely to leave their employers. These tendencies appeared to increase with plan generosity.

[24] See Luther Tweeten, "Job Search Behavior and Its Impact on Earnings," in ibid., vol. 3, pt. 2, chap. 7; and Richard E. Miller, "Job Change Behavior," in ibid., chap. 8.

Individuals in experimental groups who left their initial employer were unemployed about three weeks more than those in control groups over a two-year period. Unemployment duration for members of experimental groups compared to control groups was greater for those who had low wage earnings prior to the experiment, faced high implicit tax rates, or had incomes from another worker in the family. Younger and better educated experimental members were also unemployed longer.

Experimental individuals who changed jobs tended to obtain jobs with more desirable nonwage characteristics than similar people in control groups if they had relatively desirable jobs to begin with, and to do worse than those in control groups if they initially had relatively undesirable jobs. Experimental heads with secondary earners in their families were able to obtain jobs with better earnings prospects and higher status, particularly if they were on high guarantee plans and the secondary earner had relatively high earnings.

A standard experimental plan (50 percent tax rate, basic benefit level of 75 percent of the poverty line) led to increased wages in subsequent jobs, presumably due to a longer job search. But the increases were not statistically significant, and the gain in earnings in one year was more than offset by the earnings lost while unemployed. Higher tax rates and guarantee levels, however, significantly reduced wage gains, increased unemployment, and reduced earnings.

An analysis of job search methods showed the U.S. Employment Service to be most effective, far exceeding all other approaches. The infrequency of its use suggests that rural families in the selected areas had inadequate access to it.

Geographic mobility [25].—Families in the treatment groups had a 17 percent higher incidence of residential change than similar control families. This differential was statistically significant at the 0.05 level for families with a wage earner at the beginning of the experiment. Essentially all of this movement occurred among families in North Carolina.

Psychological well-being [26].—Scales were constructed to measure a variety of aspects of mental health and psychological well-being—for example, self-esteem, psychosomatic and nervous symptoms, positive and negative emotional states, life satisfaction, a sense of powerlessness, and a sense of being cast adrift in a chaotic world (anomie). A number of single-item questions were asked each participant—what were his or her worries over money, health, jobs, and other problems; feelings about the quality of life; hopes and aspirations for the future; subjective sense of general health; and so forth.

There was no consistent pattern in these individual measures of well-being. While some statistically significant effects were found, they were scattered and unstable over time. An overall index of psychological well-being formed from a number of individual scores showed a more consistent pattern, however. For both adults and teenagers, the three most generous plans tended to produce higher scores on this index than those of comparable control families, with the differentials statistically significant for three of the six plan/age groups. The least

[25] See Aaron C. Johnson, Jr., "Geographic Mobility," in ibid., vol. 6, pt. 1, chap. 9.
[26] See Russell Middleton, "Psychological Well-Being," in ibid., chap. 7.

generous plans tended to have a negative differential in the well-being index—significantly so for adults on the least generous plan—so that the overall experimental/control differential was small. Similarly, the level of the guarantee had a significant positive effect on the overall index of well-being for both adults and teenagers.

Marital dissolution and family interaction [27].—Income maintenance payments have an ambiguous effect on marital stability: they might strengthen the relationship by raising family income or, alternatively, they might facilitate marital dissolution by providing an alternative source of income to wives. Overall the treatment group appeared to have a higher incidence of divorce, separation, and desertion than similar control families, with families in the least generous plans showing the highest rates of dissolution and those in the most generous plan showing lower rates than control families. These differentials were generally not significant; however, the level of the guarantee was found to have a significant negative effect on the rate of dissolution.

No significant treatment effects were found for a variety of measures of family interaction, including marital happiness, satisfaction in marriage, marital disagreements, parent-child relations, and the division of household tasks within the family.

Aspirations, school attitudes, school behavior, and delinquency among teenage youth [28].—Self-administered questionnaires dealing with a variety of attitudes and behaviors were completed by 445 youths fourteen through eighteen years of age at the end of the experiment. In general, no systematic significant overall differences were found between youths in treatment families and those in controls. Areas analyzed included educational and occupational aspirations and expectations, self-rating of school ability, self-reports of grades and school behavior, interest in grades, hours of homework, participation in extracurricular activities, and general attitudes toward school and teachers. Moreover, although a relatively high incidence of delinquent behavior was reported, no significant overall differences were found between treatment and control groups under any of several measures of delinquency analyzed. The level of the guarantee did, however, have a significant negative effect on delinquent behavior; this effect was offset by an equally significant positive effect common to all plans, so that only in the plan with the highest guarantee was the delinquency rate lower than that of youths in control families.

School performance [29].—Four measures of school performance—attendance, comportment, academic grades, and standardized achievement test scores—were analyzed for 847 children who were in school at the beginning of the experiment and had completed at least one year of school after that. The treatment effects for the four age/site groups analyzed are shown in table 5. The experiment significantly improved performance by elementary students in North Carolina by all four measures, although there were no statistically significant effects among the other three groups.

---

[27] See Russell Middleton and Linda Haas, "Marital Dissolution and Family Interaction," in ibid., chap. 8.

[28] See Russell Middleton, Linda Haas, and Ain Haas, "Aspirations, School Attitudes and School Behavior of Teen-age Youth," and Middleton and Ain Haas, "Delinquency of Teen-age Youth," both in ibid., vol. 6, pt. 2, chaps. 10 and 11, respectively.

[29] See Rebecca Maynard and David L. Crawford, "School Performance," in ibid., chap 12.

TABLE 5.—EFFECT ON SCHOOL PERFORMANCE OF THE RURAL INCOME MAINTENANCE EXPERIMENT, BY SITE

[Difference between experimental and control means as percent of control mean]

Performance measure	Grades 2–8		Grades 9–12	
	North Carolina	Iowa	North Carolina	Iowa
Days absent	[1]—30	—20	3	—17
Comportment	[1]7	0	[2]	[2]
Academic grades	[3]6	—5	4	—5
Achievement tests (deviation from norm)	[1] 19	—18.8	[2]	[2]

[1] Significant at the 0.05 level.
[2] N.E.—Not estimated, due to lack of data.
[3] Significant at the 0.10 level.

Source: Rebecca Maynard and David L. Crawford, "School Performance," in *Rural Income Maintenance Experiment*, vol. 6, pt. 2, ch. 12, p. 38.

The differential strength of the treatment effect in North Carolina as compared to Iowa, may reflect the lower socioeconomic status of families and the lower educational achievement of children in North Carolina at the beginning of the experiment. For example, about 62 percent of the North Carolina families were below the poverty line initially, as compared with 37 percent in Iowa, and the mean ranking of North Carolina children on nationally standardized tests was in the 25th percentile, as compared with the 50th percentile for Iowa children.

Political participation [30].—Two measures of political participation, voting and an index of interest and involvement in political campaigns, were analyzed for the 1970 and 1972 elections. Separate analyses were performed for husbands and wives in two-parent families and for female heads of families. The treatment effect was positive in all twelve cases, although only significantly so for wives, whose voting probabilities in the two elections were increased by 6 to 10 percentage points.

ADMINISTRATION OF A NEGATIVE INCOME TAX IN RURAL AREAS

Self-employed farmers present unique difficulties in the administration of a negative income tax because their income flows are much more complex and irregular than those of wage earners. Although administrative rules and procedures were not varied experimentally, the experiment provided a good deal of practical experience in the design and implementation of rules governing the treatment of income, expenses, and assets of self-employed farmers as well as information about participant understanding of, and compliance with, these rules. The rural experiment represents the only systematic attempt to date to deal with these issues, even though farmers may

[30] See Joseph Heffernan, "Political Behavior," in ibid., chap. 13.

comprise as much as 18 percent of all male heads eligible for assistance under a universal income maintenance program.[31]

The experiment established rules for the definition of self-employment and developed a method of calculating income for the purposes of a cash transfer program. This accounting method differed from the Internal Revenue Service rules in disallowing accelerated depreciation and the investment tax credit, adding the value of rent-free housing to income, and imputing to income a percentage of assets above a given level. A one-month accounting period with a twelve-month carry-over provision was developed to deal with the seasonal variability of farm income. Experience in administering the program led to additional recommendations requiring the accrual method of accounting rather than the cash method, and treating both realized and unrealized capital gains as income.

Participants' understanding of the experimental rules was very poor.[32] Only about one-half of the families understood the guarantee, tax rate, and breakeven level they faced, and understanding of these program characteristics did not improve despite the careful instruction of participants. Nevertheless, as we have seen, there were definite changes in labor supply as a result of the experiment.

Data on family size, wage income, and transfer income were reported with acceptable accuracy, but assets were underreported on the monthly report forms by 14 to 27 percent, and farm income by 39 percent, as compared with the edited interview data.[33] On the basis of these results, underreporting by farmers could be expected to affect program costs far more than any likely change in their labor supply.

SUMMARY

Many of the results of the rural income maintenance experiment closely resemble the results of the New Jersey experiment. In families of wage earners in the rural experiment there was a somewhat larger decline in income relative to that of control families, than there had been in New Jersey, but the decline was still modest. In the rural experiment husbands' hours did not show a consistent decline, and those declines that were found tended to be even smaller, on the average, than in New Jersey. As in New Jersey, husbands did not withdraw from the labor force, but the percentage of working wives fell considerably. One new outcome in the rural experiment was that wage work of dependents also fell. But since wives and dependents had worked only a small number of hours before the start of the experiment, the effect on total family work for wages was small.

Among farm families, there was a marked reduction in hours of work for wages, with a reduction for husbands of 10 percent in Iowa and 31 percent in North Carolina, and wives working 50 to 60 percent fewer hours than those in the control group. Hours of farm work reported by farm operators rose by about 11 percent in both sites, however, so that total hours of work by farm operators and their wives actually rose in Iowa. When combined with the observed decline in

---

[31] *Social Security Amendments of 1971*, Report of the House Committee on Ways and Means on H.R. 1, H. Rept. 92-231, 92:1 (GPO, 1971), p. 230.

[32] See William S. Harrar, "Participants' Understanding of the Experimental Program," in *Rural Income Maintenance Experiment*, vol. 2, chap. 5.

[33] See William S. Harrar, "Quality of Wage Income and Hours Data," and Harrar, "Accuracy of Self-Administered Reporting," both in ibid., chaps. 2 and 4.

net farm profits in the treatment group, this increase in farm labor supply implies a reduction in the efficiency of farm operations.

As in the New Jersey experiment, there was a negligible overall effect in terms of various psychological variables, but there were some nonlabor market consequences of note. Of particular interest were the relative improvements in nutrition and in school performance of grade school children among North Carolina experimental families. A positive experimental effect also occurred for several forms of consumption, including purchase of cars, durable goods, and houses.

The experiment also indicates the need for special care in defining administrative and reporting procedures for self-employed farmers in order to avoid serious problems of underreporting and misreporting of income and assets. An accurate measurement of farm income and assets may be of greater importance among this population than any likely labor supply response.

Summary of Conference Discussion

(By Marvin M. Smith)

The preceding chapters provided the background for a two-day conference on the evaluation of the rural income maintenance experiment held at the Brookings Institution on January 13 and 14, 1977. The conference afforded the opportunity not only to evaluate the results of the rural experiment, but also to compare these results with those from the New Jersey experiment [1] and to contrast the operational design of the rural experiment with those of the ongoing Gary and Seattle-Denver experiments.

The conference was marked by a spirited debate on a wide range of issues that encompassed the theoretical difficulties that inevitably beset such a large-scale social experiment, as well as specific operational concerns encountered in its implementation, and, of course, the policy implications. This chapter presents the highlights of the conference discussion.[2]

EXPERIMENTAL DESIGN: A RETROSPECTIVE VIEW

The paper by Harold W. Watts and D. Lee Bawden provided a review and assessment of the consequences of the basic decisions and omissions that shaped the design of both the rural and New Jersey negative income tax experiments. The authors distinguished between two types of issues: those anticipated and dealt with given the best information available and those which they referred to as "surprises," and which might have led to different results had they been anticipated in advance. Included in the former category were the variety of experimental treatments, the sample selection strategy, and the time-horizon issue; while the latter issues concerned the scarcity of *very* poor families eligible for inclusion in the sample, and the altering of the New Jersey welfare program during the urban experiment.

---

[1] See Joseph A. Pechman and P. Michael Timpane, eds., *Work Incentives and Income Guarantees: The New Jersey Negative Income Tax Experiment* (Brookings Institution, 1975).

[2] Whenever the phrase "general (dis)agreement" is used in the summary, it should be noted that: (1) this does not indicate unanimous (dis)agreement on the part of the conferees, but merely suggests that a majority were in (dis)accord; and (2) no formal consensus was solicited, rather the judgment is based on the author's appraisal of the discussion.

The Guarantees and Tax Rates

In their discussion of the guarantee and tax rates employed in both the urban and rural experiments, Watts and Bawden raised a number of points concerning the variety and range of the experimental treatments. Rather than having one guarantee-tax rate scheme (and a control group) or limiting the treatment to variations of only one or the other of those parameters, the authors clearly endorsed the strategy of applying a variety of plans that varied both (as was done in the rural and urban experiments). There was no disagreement voiced on this point by any of the conference participants.

Conference discussion of the paper initially focused on the range of the experimental treatments. In their paper, Watts and Bawden pointed out that very few families were available for analysis in the 70 percent tax rate (the highest) in both the urban and rural experiments. As a consequence, the two experiments provided little evidence for assessing the possible labor supply response of families confronted with tax rates above 50 percent. They cited the confounding effects generated by the aid to families with dependent children—unemployed parents (AFDC-UP) program in New Jersey [3]—a welfare program initiated in the state after the urban experiment was under way that covered the same population—as a factor in the urban experiment, and the low policy importance attached to the 70 percent tax rate as a contributing factor in the rural experiment (the latter led to low policy weights and few observations assigned to the 70 percent tax rate cells in the sample allocation process). They further noted that the low breakeven levels of income—the income levels at which payments fall to zero—associated with such a high tax rate resulted in some families being ineligible for payments during the experiment. Although it was generally agreed that these suggested explanations had merit, several conferees sought to explore their implications and to offer alternative causes.

One of the formal discussants suggested that the design of the Conlisk-Watts sample allocation model [4] may have led to the relatively small number of families allocated to the 70 percent tax rate cell. He felt that some consideration should be given to the appropriateness of the model by reassessing its underlying assumptions. While some participants acknowledged that the allocation model may have contributed to the shortfall of families in the 70 percent tax rate cell, there were considerable differences of opinion over its relative importance—especially given the confounding effects of the AFDC–UP program on the urban experiment's results.

---

[3] The welfare payements in the generous AFDC UP plan tended to dominate the experimental payments in two of the least generous experimental plans—those with a 50 percent guarantee and a 50 percent tax rate (the 50–50 plan) and a 75 percent guarantee and a 70 percent tax rate (the 75–70 plan). Since families were given the opportunity to choose between payments from the welfare program or the experiment, many opted for the welfare payments. This attrition significantly reduced the overall sample size, particularly in the 70 percent tax rate cell.

For a detailed discussion of this point, see Pechman and Timpane, eds., *Work Incentives and Income Guarantees*, especially the formal discussion by Henry J. Aaron, "Cautionary Notes on the Experiment," pp. 88–110.

[4] For a description of this model, see John Conlisk and Harold Watts, "A Model for Optimizing Experimental Designs for Estimating Response Surfaces," in American Statistical Association, *Proceedings of the Social Statistics Section* (1969), pp. 150–56. Further discussion of the model can be found in Charles E. Metcalf, "Sample Design and the Use of Experimental Data," in Harold W. Watts and Albert Rees, eds., *Final Report of the New Jersey Graduated Work Incentive Experiment* (University of Wisconsin-Madison, Institute for Research on Poverty, and Mathematica, 1974), vol. 2, pt. C, chap. 5. The use of the model enabled the experimenters to make more precise estimates of experimental effects given the budgetary constraints and the desired stratification by age and sex of household head.

In discussing the implications of the breakeven levels of income, some participants argued that the truncation of the sample (there was a requirement that families have incomes in the year prior to the experiment of 150 percent or less of the official poverty line in order to be eligible for inclusion in the sample) excluded most families with multiple earners, thus precluding valuable information on their behavioral and attitudinal responses. One conferee added that the fact that the marginal tax rate changes abruptly at the breakeven point, when coupled with the truncation problems, makes the generalization to a national program very difficult. He felt this to be a potentially crucial problem because a national negative income tax program, if instituted, would have to be merged in some continuous manner with the positive income tax.

On the same subject, another formal discussant cautioned that the sample truncation in the rural experiment may not only have yielded fewer secondary workers in the sample, but may also have caused additional estimation problems. He pointed out that recent methodological research by Hausman and Wise [5] indicates that even when considering only single-earner families, the act of sample truncation creates numerous statistical problems for nonexperimental uses of the data. While this point was noted by the participants, it was not pursued any further.

Sample Selection

In both the urban and rural experiments, a few experimental sites were carefully selected in a nonrandom manner. This type of strategy involved the studying of households "in specific localities that were thought representative of major and possibly distinct types of labor markets." An alternative option would have been to utilize the "national sample" strategy—which entails a highly dispersed but randomly selected sample "that could be readily and routinely generalized to all geographic areas or a wider range of household categories." Further, the scattered sample approach employed in the two experiments might be contrasted with the "saturation" approach—in which "all families in one or several areas are eligible for an experimental treatment." [6] After reviewing the objectives of the two experiments, as well as considering the administrative feasibility and budgetary implications of the aforementioned strategies, Bawden and Watts maintained that the scattered sample approach of predetermined sites remained the most appropriate choice. While most conferees agreed with the wisdom of this decision, a number of points were raised concerning the relative merits of the various sample strategies.

One participant thought a strong case could be made for pursuing the "saturation" approach. He argued that when only a small percentage of the people in a given locality are involved in an experiment, the results obtained might differ considerably from those resulting from a national program which would affect *all* the low-income people in the area. These possible "community effects" on the tastes, attitudes, and behaviors of the people as well as employers, he contended, may

[5] Jerry A. Hausman and David A. Wise, "The Evaluation of Results From Truncated Samples: The New Jersey Income Maintenance Experiment," *Annals of Economic and Social Measurement*, vol. 5 (Fall 1976), pp. 421–45.

[6] The quotations are from Harold W. Watts and D. Lee Bawden, this volume, p. 58.

not be inconsequential. While recognizing the desire to have a more representative national sample, he felt that the potential worth of saturation strategies should be reconsidered.[7]

Another conferee, concerned with the large outlay and effort expended on income maintenance experimentation covering four separate experiments, suggested that, in retrospect, a national sample could have been designed that led to well-structured site-specific experiments. These experiments, he argued, would have yielded a more representative distribution of the nation's population. In responding to this point, Watts concurred that the notion of planning and administering one large experiment in lieu of several smaller ones seemed quite appealing. However, he felt that the planning for such an endeavor—had that option been chosen—would be so time-consuming that the experiment might still be in the design stage.

Duration of the Experiment

The rural and urban (New Jersey) experiments were each conducted for a three-year period. The significance of the duration of the experiments becomes quite clear when one considers whether the behavior exhibited on the part of the experimental groups (for example, the labor supply response) simulates the long-run effects that would prevail under a permanent negative income tax or merely represents temporary responses to a short-term program. The conference participants expressed divergent viewpoints on the appropriateness of a three-year time span for the experiments. One group supported the basic arguments presented in the conference paper, namely, that extending the duration of the experiments, while desirable on some grounds, would have served to (1) postpone the availability of the final results, and (2) increase costs, which, in turn, would have reduced sample sizes (given a fixed budget).

Another group argued that the continued uncertainty about the validity of the results from a short-term experiment for a permanent program suggests that either the time span of the experiments should be increased or the duration made an explicit experimental variable. In particular, one of the formal discussants found the argument of "an increased waiting time for results" to be somewhat questionable. He pointed out that a multiple duration experiment generates behavioral results in the early stages as well so that during the first three years of the experiment, families on five- or twenty-year plans, for example, should be exhibiting differential behavior.

Commenting on the budgetary concerns of a variable duration experiment, one conferee suggested that the problem may not be with the level of the budget, but rather with its allocation among different periods of treatment. He contended that the same budget would buy approximately an equivalent number of family years of information if allocated over three-, five-, or twenty-year payments. The most noteworthy difference would be that the information would be coming from different families.

After much discussion, the conference participants remained divided on the time-horizon issue. It was pointed out, however, that the design of the Seattle-Denver experiment does in fact introduce the

[7] For a discussion of saturation experiments and the possible problems involved in carrying them out, see Larry L. Orr, Robinson G. Hollister, and Myron J. Lefcowitz, eds., *Income Maintenance: Interdisciplinary Approaches to Research* (Markham, 1971).

duration of experimental treatments as a variable—with most families receiving payments for three years, some for five, and a few for twenty. The hope was expressed that the results from this experiment would provide more definitive information on this score.

Where Are the Poor?

What would a study of poor people be like without an abundance of poor people? One of the unexpected surprises in the urban experiment and the Iowa site of the rural experiment, according to Watts and Bawden, was the scarcity of families meeting the income eligibility criteria for inclusion in the samples. In the urban experiment, the authors noted that "using the income criterion of 150 percent of poverty, we found a lower density of eligible families in designated poverty areas than census data suggested: more than two-thirds were above the poverty threshold, and virtually none were below 75 percent." As far as the rural experiment was concerned, they were equally surprised to find "farmers with low incomes and low equity in their business assets, but with large gross incomes, in control of sizable business assets, and living in a style comparable to middle class families." [8] Given such experiences, the authors were prompted to ask whether the poor really existed in the numbers that the Bureau of the Census figures suggested.

This inquiry drew quick responses from some of the conference participants. Most of the comments focused on the nature of Census data. One conferee thought the Census income figures used in identifying the poor are themselves suspect. According to him, there are differential underreporting effects in the Census tabulations by source of income. He further noted that a close examination of the amount of transfer payments reported in the Bureau's field surveys indicates that public assistance, for example, is underreported by nearly one-fourth to one-third.[9] As a consequence, a large amount of income in the lower 10 to 20 percent of the population simply goes unreported. In addition, he argued that any reasonable attempt to adjust the Census figures may result in merely placing individuals into higher income brackets so that "the alleged poor simply are not there."

Another participant observed that the Census projections contain very few able-bodied, prime-age male heads of households—the prime target group of the experiment—who have poverty incomes that are substantially below the poverty level. Most of them, he noted, have incomes close to the poverty line. So the fact of not finding a large number from this group, at least, is not so surprising.

Those conferees who commented on the whereabouts of the poor seemed to agree that much more accurate information is needed about the so-called poor population in order to obtain better estimates of the number of families that might be eligible for a negative income tax; but they were confident that the poor were there.

FARMERS' LABOR SUPPLY RESPONSE

In his paper assessing the labor supply responses of farmers in the rural experiment, Finis Welch did not present a critique of the estimates of farmer labor supply obtained by the rural experiment's ana-

[8] The quotations are from Watts and Bawden, pp. 63, 64.

[9] For a study documenting this occurrence, see Joseph J. Minarik, "New Evidence on the Poverty Count," in American Statistical Association, *Proceedings of the Social Statistics Section* (1975), pp. 554–59.

lysts, but opted instead to "spell out why we should have little interest in these estimates." [10]

Using data from the rural experiment, Welch performed some calculations of his own. In one such set of calculations, he employed a three-way split of the combined data from Iowa and North Carolina—namely, females, males who worked off the farm in 1969, and males who did not work off the farm in 1969—to obtain "averages" of hours worked on and off farms for experimental and control farmers along with net farm income for the three years of the experiment. After cautioning that the "averages" obtained in his analysis should not be construed as accurate estimates of program effects, but instead viewed as indicative of general response patterns, Welch showed that total hours of work by the various groups changed very little. The composition of hours worked, however, changed dramatically. There was significant substitution from off-farm work to farm work. Coincidentally with this compositional change, reported farm income decreased noticeably. The author suggested as a possible explanation that only reported and not actual farm income fell for these groups, stating that "the overwhelming evidence of the analysis file is that farmers underreport income." Welch pointed out that there was a marked difference between the stated and actual tax rates faced by the treatment group—for example, a dollar of net farm income added 9 cents to taxable income in the range of taxes that included 70 percent of the participants; thus, there existed approximately a 3 to 6 percent tax rate on farm income.[11]

As a result of what he believed to be shortcomings in the experimental design (some of which are discussed below), Welch viewed the quality of the data with suspicion and thus was not inclined to place much faith in any of the results reported. During the discussion, he reaffirmed his skeptical view by noting that even though one might agree that complicating factors and unforeseen events would no doubt accompany a real world negative income tax program, it is questionable that we can simulate an average amount of such distortions in a three-year experimental program. Therefore, even though the experimental results may indicate little labor supply effect, it is debatable to conclude that there is likely to be little effect in a permanent income maintenance program. As far as he was concerned, the experimental program implemented in Iowa and North Carolina was not a textbook negative income tax but something quite different.

Welch's views provoked spirited and wide-ranging responses from the conference participants. These are discussed below under four categories: (1) estimation and related problems, (2) accounting procedures, (3) underreporting of income, and (4) general problems.

Estimation

It was generally agreed that the small sample size for farmers weakened the power of the statistical tests of significance. However, some conferees questioned the wisdom of Welch's decision to aggregate experimental groups across plans and states in carrying out his analysis. Although it was done presumably to ensure the reliability of the results, some thought that the procedure combined reactions in two

[10] This volume, p. 99.

[11] These estimates are based on analysis of benefits in relation to the experimenter's best estimate of "true" farm income.

sites that were structurally different and thus precluded the analysis of site-specific differences in the employment and earnings of those in experimental groups.

One of the formal discussants noted that the sample size problem might have been circumvented had the author taken into consideration that a pooled analysis could have been undertaken with the appropriate use of dummy variables to identify the two sites—which, in addition, would preclude the performing of separate analysis for each state.

There was also some concern expressed that experimental families with incomes either negative or above the breakeven level might have faced different real tax rates from those confronting other experimental families assigned to the same plan. The apprehension was that since experimental families with earnings above breakeven levels were confronted with tax rates similar to those faced by control families, their behavior might be more like that of the controls. Although many of the conference participants joined Welch in acknowledging the problematic nature of this situation, some of the participants not only viewed the problem as a tractable one, but also felt the issue was somewhat exaggerated in the paper. While there was no consensus on a solution to this "breakeven issue," one conferee suggested that the problem could be handled by simply having higher guarantees or having only families with lower incomes, thereby ensuring that none of the experimental families was above the breakeven level.

Accounting Procedures

A number of problematic accounting features were addressed in the paper and debated at the conference. One problem noted in the discussion was the use of cash basis accounting in the computation of income for determining benefit payments. Under this method, costs incurred and revenues received are accounted for only when they are realized and not as they accrue. Further, the income upon which payments were made was computed on the basis of a three-month moving average, with a twelve-month carry-over of excess income or expenses to be counted in the three-month average. According to Welch, this procedure probably led to questionable behavior by some farmers, thus confounding the interpretation of the results obtained. He illustrated his point with the following example. Consider a farmer who produces durable products (corn in the case of Iowa and tobacco in North Carolina) that have rather low storage costs and who is confronted with an experimental tax rate of 70 percent. There is a real question as to whether any of the products would be sold during the period of the experiment. As long as the rural experiment was willing to subsidize 70 percent of the expenditures incurred in raising the crop, the product would be held off the market until after the experiment ended so that the revenues from the sale of the crop would be taxed at the IRS norm and not at the 70 percent rate. One conferee, however, was very skeptical that such a distortion was of major consequence. He noted that on the basis of postexperimental data, there was no evidence that goods were in fact withheld from the market until after the termination of the experiment. It was pointed out that farmers cur-

rently employ cash accounting in income tax computation, and no doubt "would press to have the privilege extended in a negative income tax."

Another feature of the rural experiment's accounting system involved the double taxation of farm assets. Welch pointed out that under the "rules of operation" as set forth in the experiment, net equity in farm-owned assets (such as farmland, buildings, machines, livestock, and grain inventories) above $20,000 was taxed directly and the revenue generated from these assets was later taxed. One of the formal discussants stated that "society does not favor income transfers to households with massive wealth, even if that wealth was accumulated out of past earnings that were taxed," and therefore felt that double taxation would probably continue. While the author was sympathetic to the idea of not paying an individual welfare "if he is sitting on $100,000 worth of land," he had some reservations as to whether double taxation of farm income would attend a national negative income tax program.

The third and, according to Welch, "the most interesting, most subtle, and most important wrinkle of the rural experiments accounting" concerned the "carry-over" provision.[12] In his paper he demonstated that, with the aid of the carry-over provision and the appropriate sequencing of family income, it is possible for two families with identical incomes assigned to the same experimental treatment plan to end up three years later with different benefits received.

In response to this demonstration of the gains from sequencing under the rules of the experiment, several participants felt that even if sequencing were not available, farmers were likely to seek other legal avenues to circumvent their tax liabilities. Welch acknowledged that if a national negative income tax program were put in operation for many years, most of the problems associated with the accounting system would go away. However, he maintained that his major concern was that in an experiment of short duration the incentive structure is *very* different; therefore, it is debatable whether the near zero labor supply response under this structure would prevail in a national income maintenance program.[13]

Underreporting

Perhaps the most debated issue was the underreporting of farm income for payment purposes. There was much discussion about its origin and significance with most conferees agreeing that underreporting could have compromised the quality of responses obtained during the quarterly interviews.[14]

One participant suggested that the short-term nature of the rural experiment may have generated behavioral responses such as the sequencing of income, which altered the pattern of income flows and,

---

[12] The carry-over provision was an experimental design feature to take account of the uneven flows of farm income throughout the year. Earned income in excess of the breakeven level was carried forward for a maximum of one year and added to income in any period in which it fell below the breakeven level. Benefit payments were calculated on the basis of the sum of current income and the carry-over assigned to that period.

[13] It was pointed out, however, that since preenrollment income entered the calculation for the first two payments, Welch's example of what might happen may be somewhat exaggerated. It was further noted that even in his extreme example, equilibrium payments would be reached before the end of the experiment.

[14] In fact, the quarterly interviews on which analyses of behavioral responses were based were first edited to remove a substantial portion of the underreporting.

in turn, resulted in the actual levels of reported income to be lower than expected by the experimenters.

However, to the extent that there was some underreporting, others argued that the misreporting of income by farmers is a problem that the Internal Revenue Service has and will continue to have, and one that any national income maintenance program will have; therefore, one should not lay the underreporting issue at the doorstep of the experiment. If anything, it was contended, the rural experiment simply replicated what should be expected in an actual program.

There was considerable disagreement among the conference participants on the extent of the underreporting that took place. One conferee, citing a study he had completed on the subject, indicated that in terms of cash income and expenses, farmers on the average reported only 60 percent of what the experimenters thought their real yearly income to be—excluding capital gains and depreciation. Further, underreporting seemed to be concentrated in gross omissions of gross income—as opposed to padding expenses. In addition, farmers underreported their assets—principally farm machinery and land—by between 15 and 25 percent (see the paper by Bawden and Harrar in this volume). Some participants questioned the validity of the information since it hinged upon knowing what the farmers' true incomes were. This point, as well as the general issue of underreporting, was debated at length but there was no resolution of the issue.

General Problems

The finding of a small overall change in the labor supply response of farmers—and its use as an indicator of the relevant response in a national program—not only bothered Welch but concerned other conference participants as well. One conferee stated that since there are so many potential biases in the data, he was unwilling to conjecture whether there would be a large or small labor response in a national program. He added that, in all probability, any experiment of short duration would grossly underestimate the true income effect on labor supply and no doubt overestimate the true substitution effect.[15]

One curious finding noted in the conference paper was that farmers tended to decrease their wage work and increase their farm work while their farm profits decreased. Some participants suggested that this may have resulted from an increase in the consumption value of farming (via income in-kind). In short, the true labor supply function should not depend upon net wages and net income only but should also take into consideration the price of other commodities.

In view of the many objections concerning the origin and interpretation of the labor supply responses of farmers, some conferees suggested that labor supply, as measured in hours, may not be the proper issue. Instead, it was argued that the appropriate issue should be the "work effort," as measured in output, expended by farmers. As one participant suggested, more attention should be focused on the output from farms—which, incidentally, declined very slightly. Still others felt that more concern should be given to the income of farmers rather than their hours of work, as an indicator of the well-being of farm families.

---

[15] This observation was in fact verified in the New Jersey experiment. For a discussion, see Charles E. Metcalf, "Predicting the Effects of Permanent Programs from a Limited Duration Experiment," in Watts and Rees, eds., *Final Report of the New Jersey Graduated Work Incentive Experiment*, vol. 2, pt. C, chap. 3.

Perhaps one of the overriding issues in the discussion was the extent to which the rural experiment demonstrated how a negative income tax program would actually operate in the real world. Unlike the author, many conferees felt that it was a real world demonstration despite the confounding design features and the unforeseen behavioral responses on the part of farmers. The significance of the experiment in their viewpoint was perhaps best summed up by Luther Tweeten who stated in his formal comments that, "If the experiment is in fact a demonstration of what we might expect from a negative income tax scheme actually implemented, it is important to examine the results for what they can tell us about the real world."

WAGE EARNERS' LABOR SUPPLY RESPONSE

According to Orley Ashenfelter, the author of the conference paper on the labor supply responses of rural nonfarm wage earners, "The results of the rural negative income tax experiment show an unambiguous average decline in the work effort of all family members in the experimental group of wage earners." He arrived at this conclusion after reviewing the results obtained by the rural experiment's analysts as well as his own estimates derived from the experimental data. Ashenfelter's calculations showed that the wage income of husbands declined by 8 percent and that of wives by 27 percent, while family wage income (the weighted sum of the results for husbands and wives) averaged a 12 percent decline. These estimates, as noted by the author, are quite close to the regression estimates computed by the experiment's analysts, except for a discrepancy in the experimental effect on total family wage income—which can be attributed, at least in part, to the presence of additional wage earners in the analysts' family estimates. In addition, these results from the rural experiment, though somewhat smaller, closely resemble the estimates calculated by Robert Hall of the urban experiment's effect.[16] However, Ashenfelter points out that since there are differing estimates of the effect of the urban experiment on labor supply,[17] the comparability of the results between the two experiments depends on whose urban estimates are used.

Nonetheless, the results of the rural experiment drew mixed responses from the conference participants—some tending to accept them as reasonable estimates, with others placing less faith in them. During the course of the debate concerning these results, two topics tended to dominate the discussion: (1) the experimental treatments, and (2) the ramifications of program costs.

Experimental Treatments

A number of conferees voiced concern about the proposition that the implicit tax rate in most welfare programs (including the rural experiment) is not in fact the statutory tax rate. Ashenfelter suggested that this possibility might be illustrated within the context of a pure tax avoidance model. He proceeded by noting that, on a pure tax avoidance basis, one would expect people in a negative income

[16] See Robert E. Hall "Effects of the Experimental Negative Income Tax on Labor Supply," in Pechman and Timpane, eds., *Work Incentives and Income Guarantees*, pp. 115-47.

[17] See Aaron, "Cautionary Notes on the Experiment," and Albert Rees and Harold W. Watts "An Overview of the Labor Supply Results," in Pechman and Timpane, eds., *Work Incentives and Income Guarantees.*

tax program to report less earned income simply because that would raise their payments—as indeed occurred in the rural experiment. An interesting question then becomes whether a change in reported income reflects an actual reduction in earnings or simply less accurate reporting of the same level of earnings. According to Ashenfelter, one way to approach this issue is to consider whether tax avoidance varies with the guarantee or tax rate. He argued that tax avoidance probably does not vary with changes in the guarantees because, once they are fixed, those with higher guarantees have no more incentive to avoid additional taxes than those with lower guarantees. The tax rate, on the other hand, fosters tax avoidance since the extra amount received from a change in reported income varies directly with the tax rate. Since the experimental data showed no effect for the guarantee and large effects for the tax rate, Ashenfelter suggested that a plausible interpretation might be that there was no actual change in labor supply or earnings but simply a change in the amount of earnings reported.

There was no disagreement among the conferees on the validity of the author's reasoning concerning tax avoidance; however, some participants doubted that this was the appropriate interpretation of the results—and therefore did not accept the conclusion that the labor supply response was necessarily small.

A closely related issue involved the appropriate statistical representation of the experimental treatments. Some participants argued that the use of a dummy variable—in addition to tax and guarantee variables—to indicate whether a family was in the experiment or not may have been an unwise decision. They felt that much more work was needed on the proper representation of treatments so that more confidence could be placed in the behavioral effects of the tax rates as well as guarantees.[18]

Speaking in defense of the treatment dummy, one participant pointed out that if being in the experiment itself had some independent effect on behavior—and the subsequent significant coefficient on the treatment dummy seems to indicate as much—a major concern should be the marginal effects that the guarantee and the tax rate had beyond the effect of being in the program. Thus, he argued, the planners were correct in thinking that it would be better to use a treatment dummy to capture this independent effect than to leave it out and force all of the response through the tax rate and guarantee.

The author of the conference paper expressed some disappointment that the rural experiment, like its urban predecessor, failed to show—with any degree of statistical significance—differences in response among guarantee levels. He felt that the seemingly large coefficients for the tax rate and the rather small guarantee coefficients were indicative of the failure to capture these effects appropriately. He suspected this might have been due to the inclusion of the treatment dummy, or to not having used some other representation of the experimental treatments that would have captured the "right" (or at least a more "reasonable") response. As a result, Ashenfelter and several other conference participants questioned the validity and usefulness of the responses obtained in the rural experiment.

---

[18] One conferee did note, however, that the analysts experimented with a number of different specifications of the treatment but reported only the results yielded by the one they believed best.

A number of conferees, however, took exception to this viewpoint. Their view was best summed up by one of the participants who felt that although being able to separate out income and substitution effects would undoubtedly increase the usefulness of the results,[19] the labor supply results as a whole were of considerable value.

Program Costs

Another topic that received much discussion was the potential cost of a national program. Since one of the primary purposes of the rural experiment was to gain insight into the comparative costs for various types of negative income taxes, the conference participants attempted to focus on this objective in light of the results obtained from the experiment.

In discussing possible program costs, Ashenfelter pointed out that the labor supply figure in his analysis pertained to a program involving only families who were below the breakeven level. He noted that the inclusion of families both above and below breakeven would increase the total dollar cost of the program. This point is particularly instructive since it is possible that with a national program some families with earnings just above the breakeven point might reduce their labor supply and thus become eligible for the program, thereby increasing total program costs.

Following up this point, another participant noted that even the small labor supply responses found in the rural experiment may have significant cost implications for a national program because of the shape of the U.S. income distribution. The U.S. population has a high concentration of families near the breakeven point; these families receive relatively small transfer payments and even a very modest change in their earnings would result in a substantial percentage increase in their benefits and total program costs. This type of sensitivity of program costs to small changes in labor supply has been found in national simulations of the preliminary Seattle-Denver results and in some cross-sectional analyses.

A third participant thought the preoccupation of all the experiments with variations in the tax rate and guarantee might have been at the expense of other important variables in the design of a negative income tax program (for example, the definition of income, the income accounting period, the definition of the filing unit). Minor changes in these variables might exert a greater impact on total program costs. Further, he expressed concern that the narrow focus on these particular experimental treatments may have caused the Congress to be unresponsive to a thorough consideration of all of the elements involved in designing a national program.

In contrast, another conferee argued that it was precisely because of these experiments that so many individuals are now sensitized to the confounding nature of such design features. It was pointed out that the experiments have been extremely helpful in reducing the uncertainty of the expected labor supply response, thus enabling the policy analysts and policymakers to look to other issues. It was noted that the experiments did in fact yield considerable information on

[19] According to the conferee, the separating out of income and substitution effects could be important for (1) distinguishing between transfer costs and real resource costs, (2) simulating the costs of programs that were not tested, and (3) simulating the costs of programs for populations other than the sample populations.

some of the other design features in question, which previously had been neglected.

Other Behavioral Responses

Although the primary focus of the rural experiment was on the labor supply response to the experimental treatments, it was thought that other information on consumption behavior (for example, purchases of housing, clothing, and medical care) and noneconomic outcomes (such as impacts on marital dissolution, migration, political participation, and psychological well-being) would provide added dimension to our understanding of the total impact of a negative income tax program. This additional information might aid in the overall planning of a national program as well as illuminate other costs and benefits of such a program.

The Consumption Studies

Robert Michael, the author of the conference paper evaluating the studies on consumption behavior, summarized the findings of these papers as follows: "The additional income and the income guarantee provided by the experiment improved average nutritional intake in the North Carolina sample (but not in Iowa where the levels were initially more nearly adequate), increased consumption of clothing (somewhat less for male than for female clothing), slightly increased household inventories of durable goods and cars (except for cars for farm households), and reduced short-term farm debt and increased farm liquid assets somewhat." Although some of the preceding results closely resemble those obtained in the New Jersey experiment [20] while others reveal more consistent patterns than in New Jersey, one important caveat is in order. The implications of these observed consumption effects for a long-term income maintenance program of similar structure depend, to a large extent, on whether the experimental families regarded the payments they received as transitory income or as a permanent component of their income stream—that is, if the experimental payments were treated as permanent additions to family income, the resulting changes in consumption patterns might serve as an indication of the consumption behavior in a permanent national program.

Even though the studies on consumption behavior reported many interesting effects, Michael was somewhat disappointed in the papers' preoccupation with estimating the response to the experimental treatments—the guarantee and payment. He suggested that more descriptive studies were needed to gain insight into the expenditure behavior and mobility patterns of those "in poverty." One conference participant cautioned that in view of possible truncation of the sample, use of truncated data might lead to biased results—studies purporting to describe the expenditure patterns and poverty status of the low-income population might be grossly misleading.

At various points during the conference, participants touched on specific technical (or estimation) problems and experimental design features (such as small sample size, short duration, and availability of quality data) that plagued a number of the consumption studies. Nonetheless, the possible implications stemming from the results of the two studies concerning health effects and indebtedness received considerable attention.

---

[20] Except those results pertaining to nutrition and clothing, which were not analyzed in the New Jersey experiment.

Michael and several other conference participants were concerned that the lack of positive effects on the health variables—the conference paper reported *no* experimental effects—might be misconstrued. They feared the inference might be drawn that giving money to poor people does not improve their health and the further implication that the money was, instead, inappropriately spent by the treatment group. Pointing to such considerations as the relatively small sample size and the self-reporting problems, the concerned participants urged that the health results should be interpreted with great care.

Another thorny issue concerned the appropriate interpretation of the results on indebtedness. It was pointed out at the conference that three sources (the overview paper by Larry Orr in this volume, the original paper on this subject prepared by the rural experiment's analysts, and the summary report by the U.S. Department of Health, Education, and Welfare) all report somewhat different results. For example, one suggests loan debt substantially decreased while another indicates a substantial increase. After some discussion it was pointed out that these diverse findings could be reconciled since one source reports the long-run effect, another the short-run effect, and the third both effects. This still left unanswered the question of which results should be reported to the public at large. There was no consensus among the conference participants on this point.

One conferee observed that in spite of the many qualifications associated with the debt data from the urban, rural, and Gary experiments, one response seemed common to all three, namely, an increase in indebtedness on the part of the experimental families that was found to be tied quite explicitly to the acquisition of durable goods.

Noneconomic Outcomes

Michael Hannan's paper, which reviewed and evaluated the findings of the noneconomic outcomes of the rural experiment, reports the following results: higher overall incidence of marital dissolution among the treatment group as compared to the control group (though not statistically significant), with those on the most generous plan having lower rates (than the control group) and those on least generous plans having higher rates; a positive impact on geographic mobility (primarily in the North Carolina sample); improved performance for students in North Carolina with respect to absenteeism, comportment, academic grades, and standardized achievement test scores (with statistical significance for primary grade students); no significant overall effect on delinquency; positive treatment effect on voting and political participation (with significance only for married women); a small (and insignificant) overall treatment-control differential in the assessment of psychological well-being; and a narrowing of the occupational aspirations and expectations gap of teenagers.

To Hannan's and other conferees' surprise, many of the noneconomic outcomes in the rural experiment had positive treatment effects. However, most participants seemed to regard them as tenuous. They cited as their source of skepticism the various confounding features such as small sample size, too short a time span, low reliability of sociological and psychological variables, sample attrition, low understanding on the part of the treatment group, and the lack of well-defined theory governing the social and psychological effects studies. In fact, Hannan

suggested that the studies might be vest viewed as "exploratory." One conferee, on the other hand, aptly summed up the sentiments of many participants by stating that the rural experiment was poorly designed for the comprehensive study of sociological effects of a negative income tax.

One topic that received some attention and on which there were differing perceptions, concerned the culture of poverty theory—and what the experimental evidence suggested about poverty cycles. As Hannan pointed out, "whether or not the culture of poverty is an important causal factor in the persistence of poverty depends on whether or not individuals can shift living styles once the culture becomes maladaptive." He further noted that "if adults are indeed trapped by their origins, they cannot respond to environmental changes such as that afforded by income maintenance." Since the rural experiment allowed the income environments of families to be manipulated, Hannan felt the experiment offered a unique opportunity to separate effects of culture and class from those of current variations in employment or income." During the conference discussion, Hannan indicated that a strong reading of the culture of poverty argument suggests that no experimental effects would occur on any of the noneconomic outcomes either singularly or collectively. Whether or not the results of the experiment were inconsistent with that view depends upon how one reads the evidence. On this score, the participants were unable to reach an accord.

Hannan's reading of the results—although considered overly optimistic by some—led him to believe that there was evidence of a mild rebuff to the culture of poverty thesis. Aage Sørensen—a sociologist as is Hannan—took an opposing viewpoint. He emphasized the need to observe possible changes in the relevant noneconomic variables over several generations before assessing whether the culture of poverty thesis still holds.

OVERVIEW AND POLICY IMPLICATIONS

With the papers by Larry L. Orr, G. Edward Schuh, and D. Lee Bawden and William S. Harrar as background, the conference participants were charged with the task of making a critical assessment of the overall significance of the rural experiment and to suggest possible policy implications.

By and large, the conferees tended to fall into one of two camps as far as their general impressions of the experiment were concerned. One group viewed the experiment as providing valuable information on the labor supply response of rural families to a negative income tax and pinpointing some of the difficulties involved in administering such a program to the self-employed.

Members of the other group, while acknowledging the insight gained in administrative issues and the implications of the experiment for future social experimentation, nonetheless took a more cautious approach. Their major sources of skepticism stemmed from the problematic experimental design features and the statistical techniques used in some of the analyses.

In spite of the methodological and empirical difficulties encountered in the rural experiment, some participants thought a number of note-

worthy outcomes were obtained that had immediate policy relevance. Apart from the obvious interest in the work response of welfare recipients to governmental transfer schemes, a growing concern among policymakers is the denigrating effects associated with most public assistance programs. On this score, it was suggested in the paper by Schuh that one policy implication that might be drawn from the rural experiment results was that although "there was no evidence of an experimental effect on psychological well-being, there was also no evidence that the income maintenance program undermined self-respect or brought a reduction in self-esteem."

Another outcome that was regarded as having possible far-reaching effects from a policy standpoint was the experimental impact on political behavior. While the findings of the rural experiments indicated only a modest increase—5 to 10 percent—in the voting probabilities of those in experimental groups, Schuh suggests that this effect may considerably understate the potential impact since "the stakes of the program were relatively modest." He contends that "if the size of the income transfers were larger, there might be more substantial political participation to protect the income base, or to enlarge it." One conferee speculated that the possible fear among incumbent politicians of a *large* political impact might partially account for the difficulties being experienced by the negative income tax concept in the Congress.

A closely related issue involved the overall impact of a national program on rural towns and communities. Several participants expressed concern that the macro impact of a universal program on the social, economic, and political structure of small towns—especially in the Southeast—is very likely to generate unpredictable behavioral and structural changes that could significantly affect, inter alia, (local) labor markets, political awareness, and demographic shifts.

Commenting on the possible impact on labor markets, one conferee noted that the observed reduction in labor supply in a dispersed experiment may not be translated into an equal reduction in aggregate output. This might occur—given less than full employment—when marginal workers, responding to a negative income tax, reduce their labor supply (either partially or totally) and are replaced by others (on a part-time or full-time basis) who were nonrecipients. Since the total number of work-hours might remain approximately the same, aggregate output might show little change. Consequently, the impact might simply result in a reshuffling of unemployment with virtually on change in aggregate output.

Some participants thought that a universal program might improve the functioning of labor markets by providing recipients with the opportunity of increased job search and occupational mobility. They felt that the support payments might serve as an income cushion that could be used to defray the opportunity costs incurred when making beneficial job changes or engaging in more productive job searches. While the results from the rural experiment varied from one treatment plan to another, a standard experimental plan (50 percent tax rate and guarantee of 75 percent of poverty level) did result in members of experimental groups locating jobs with higher wages, presumably through longer job search. In addition to possible improvements from job search, Schuh points out that "another possible outcome

is that income transfers may be viewed as earning subsidies, increasing the job stability of some workers or supplementing income from jobs with good earnings prospects but low initial wages." Thus, it was concluded that although an income maintenance program, such as a negative income tax, might generate some adverse labor market effects, there are some positive outcomes affecting the job environment through more rewarding search methods that should not be overlooked.

After some discussion of the uncertainties surrounding the nature and extent of the impact of a national program on rural communities, a number of conferees were prompted to suggest that a saturation experiment—preferably in a small southeastern town—might be one possible strategy for gaining insight into the potential outcomes.

At one point in the discussion there was some concern over whether a 10–12 percent reduction in family labor supply should be regarded as large or small. Some participants argued that the labor supply response should be considered small in light of popular concern over a large-scale withdrawal from the labor force by recipients. Other conferees were much less inclined to look upon the labor supply numbers as small. Their view was best summed up by one participant who remarked that he did not regard 10–12 percent as zero. According to him, zero is a very significant political number. It was further noted by this group that the long-term effects as yet are unknown. Thus, the long-term response might be larger than the 10–12 percent obtained from the short-term experiments.

It was noted, however, that although the farm labor supply response was of central importance in the rural experiment, there are relatively few low-income, nonaged farmers in the United States. Thus, their labor supply responses—whether very large or very small—would make little difference in either the national costs of a negative income tax program or in the agricultural sector of the economy. The rural non-farm work responses, on the other hand, were considered far more important from the standpoint of a national program and were roughly consistent with the results of the New Jersey experiment.

The two days of discussion indicated that there was disagreement over the usefulness of the farm work and income responses for estimating the effects in a national program, but that the work and income responses of rural nonfarmers were viewed as useful additions to the New Jersey and cross-sectional findings for the estimation of the possible effects of a universal program. In addition most conferees generally agreed that much was learned about the administration of an income maintenance program to the rural self-employed. Some participants, in fact, argued that the rural experiment was most instructive in pinpointing and assessing the nature as well as the relative magnitude of some of the problematic areas—such as the definition of income, and the administrative and reporting procedures governing the potential underreporting and misreporting of income and assets by the self-employed—that could be expected to accompany a universal program. As Welch concludes in his paper, "when dealing with social experiments or longer-run welfare programs, it really matters how these programs are administered—more so than many of us would have suspected."

Even though the conference participants differed in their perceptions of the validity of the findings in the rural experiment, there was widespread agreement that any results reported to the public should be done with the appropriate qualifications and without bias.

35-071—78——8

IV. The Work Effort and Marital Dissolution Effects of the Seattle and Denver Income Maintenance Experiments

TESTIMONY BEFORE THE SENATE FINANCE COMMITTEE, MAY 1, 1978

(By Robert G. Spiegelman, Lyle P. Groeneveld, and Philip K. Robins)

I. INTRODUCTION

One of the primary purposes of conducting the Seattle and Denver Income Maintenance Experiments (SIME/DIME) is to provide information for the design of a national welfare program. Knowledge of the work effort and marital status effects of the experiments and the ability to extrapolate the site-specific results to the national population represent critical inputs into the overall program design. The purpose of this paper is to present an overview of the research findings from SIME/DIME with respect to marital status and work effort, and to describe how the experimental results are being used to draw inferences about the likely work effort effects of a national program.

II. A DESCRIPTION OF SIME/DIME

The Seattle and Denver Income Maintenance Experiments are testing eleven variants of a negative income tax (NIT), which is similar in structure to the cash assistance portion of the Program for Better Jobs and Income (PBJI). The experimental plans combine three support (or guarantee) levels with two tax (benefit reduction) rate systems. The three support levels (normalized for a family of four in 1971 dollars) are $3,800, $4,800, and $5,660.[1] One of the tax rate systems has constant average (and marginal) tax rates of .5 and .7. The other tax rate system has average (and marginal) tax rates that decline with income. Under the declining system the average tax rate decreases from initial values of either .7 or .8 at the rate of .025 per thousand dollars of annual income (the rate of decline of the marginal tax rate is .05 per thousand dollars of income). The experimental plans have no work requirements associated with receipt of benefits.

In order to eliminate the influence of other tax and transfer programs, SIME/DIME fully taxes public transfers and reimburses positive income taxes. A national program would presumably operate in a similar fashion. Because positive taxes are reimbursed, the payment a

[1] Adjustments are made to the support level for family size and for cost of living changes over time. In February 1978, for example, the support levels in the Denver experiment for a family of four were $5,755, $7,255, and $8,465. These support levels are substantially above the support levels of the PBJI.

person receives depends on gross income and both experimental and nonexperimental tax rates.

In Figure 1, the interrelationship between an experimental plan with a constant tax rate and the positive income tax system is depicted graphically. The horizontal axis shows gross income (income before taxes or transfer payments) and the vertical axis shows disposable income (income after taxes are subtracted and transfers are added). Two breakeven levels are distinguished. Point B is the tax breakeven level, where disposable income is equal before and after imposition of the experimental plan. All persons to the left of B (with gross income initially less than B′) are better off with the program. Point G is the grant breakeven level, the point at which the grant (payment less positive tax reimbursement) is zero. At point G an individual does not receive a grant, but also does not pay positive taxes. Table 1 presents the grant and tax breakeven levels for the eleven programs tested in SIME/DIME. As this table indicates, families not receiving grants are still eligible to receive benefits (in the form of tax relief) at fairly high levels of income.

About 4,800 families were originally enrolled in the experiments during 1971–1972. Roughly 55 percent of the families are experimental families and 45 percent are control families. About two-thirds of the experimental families are enrolled for three years while the remainder are enrolled for five years.[2] SIME/DIME also has three manpower treatments which combine job counseling with education and training subsidies.

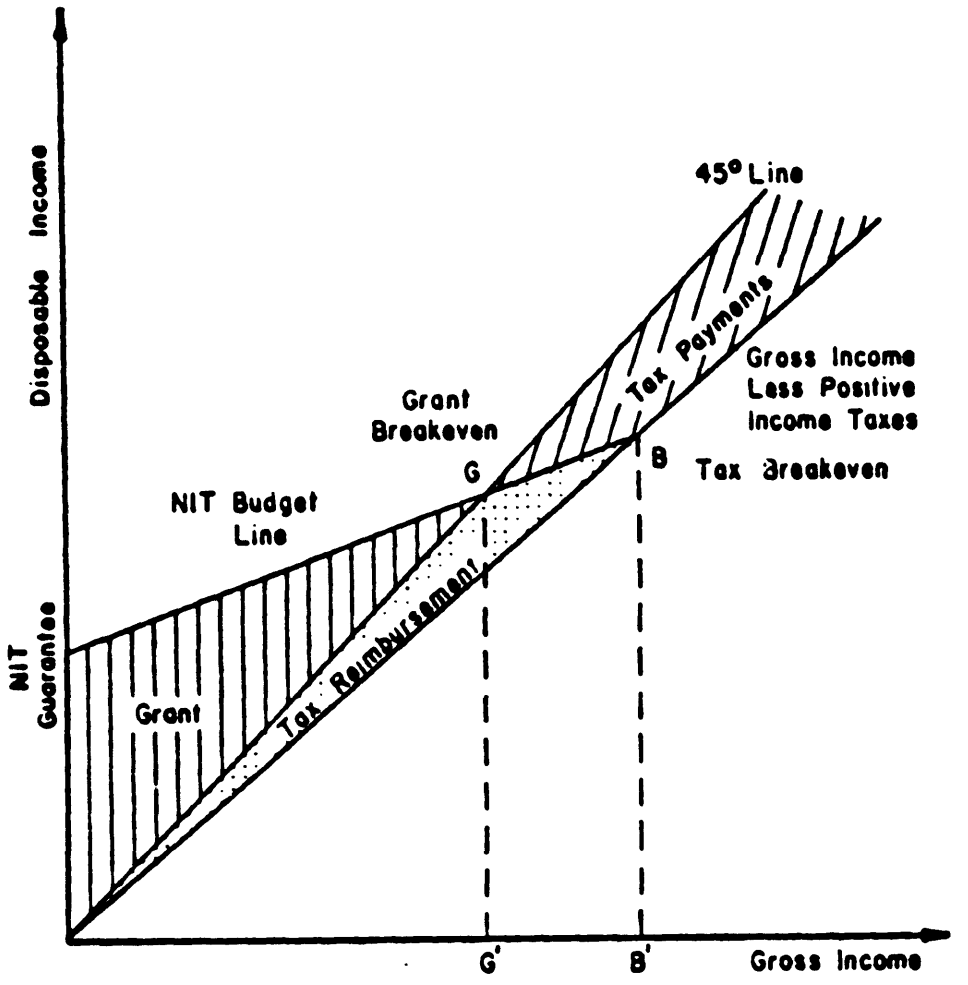
An important feature of SIME/DIME (and the other experiments as well) is a stratified allocation of families to experimental treatments on the basis of four assignment variables: family type (one or two family heads), ethnicity (Black, White, or Mexican-American), site (Seattle or Denver), and normal income (seven levels of "typical" pretransfer family income adjusted for family size).[3] The work effort results presented in this paper, cover only the Black and White family heads. Subsequent analysis has indicated that the work effort response of Mexican-American families is slightly larger than that of the other groups. The marital status results cover originally enrolled families in all three ethnic groups.

---

[2] SIME/DIME is also testing a 20-year program, which began about two years after the three and five year programs. About 170 families in Denver were assigned to treatments under this program. The work effort response of 20-year families has not yet been analyzed.

[3] To be eligible for SIME/DIME, normal income had to be less than $9,000 per year in a family of four with one working head, and less than $11,000 per year in a family of four with two working heads.

FIGURE 1.—A NEGATIVE INCOME TAX PROGRAM WITH POSITIVE TAX REIMBURSEMENTS



NOTE.—Figure assumes no income outside of earnings and a linear positive income tax system.

TABLE 1.—PLAN BREAKEVEN LEVELS FOR THE SEATTLE AND DENVER INCOME MAINTENANCE EXPERIMENTS

[1971 dollars]

Plan	Grant breakeven level	Tax breakeven level
F1 ($S = 3800$, $t_e = 0.5$, $r = 0$).....	$7,600	$10,250
F2 ($S = 3800$, $t_e = 0.7$, $r = 0$).....	5,429	6,350
F3 ($S = 3800$, $t_e = 0.7$, $r = 0.025$).....	7,367	10,850
F4 ($S = 3800$, $t_e = 0.8$, $r = 0.025$).....	5,802	7,800
F5 ($S = 4800$, $t_e = 0.5$, $r = 0$).....	9,600	13,150
F6 ($S = 4800$, $t_e = 0.7$, $r = 0$).....	6,867	8,520
F7 ($S = 4800$, $t_e = 0.7$, $r = 0.025$).....	12,000	19,700
F8 ($S = 4800$, $t_e = 0.8$, $r = 0.025$).....	8,000	11,510
F9 ($S = 5600$, $t_e = 0.5$, $r = 0$).....	11,200	15,700
F10 ($S = 5600$, $t_e = 0.7$, $r = 0$).....	8,000	9,780
F11 ($S = 5600$, $t_e = 0.8$, $r = 0.025$).....	10,360	16,230

Note: These figures are for a family of 4 with only 1 earner and no income outside of earnings. Positive tax reimbursements include the Federal income tax and social security taxes. The Federal income tax assumes the family takes the standard deduction. State income taxes, which are relevant only for the Denver Experiment (there is no State income tax in Washington), are ignored in calculating the tax breakeven levels. The tax breakeven levels are thus slightly higher for the Denver Experiment.

Key: S = NIT annual support level; t_e = initial NIT tax rate; r = rate of decline of the average NIT tax rate per thousand dollars of income (rate of decline of the marginal tax rate is $2r$).

Table 2 presents a selected number of characteristics of the Black and White families studied at enrollment. The typical sample member has income and education levels that are above the levels associated with most families in poverty. There is fairly strong attachment to the labor force among primary earners and about two-fifths of the secondary earners are employed. The sample consists primarily of young families with two children and the average initial benefit received from the experiment was about $1,300 per year which is substantially less than the average support level of $4,800 per year (the benefit received by families with no working members). Approximately 14% of the husband-wife families and 54% of female-headed families received welfare (AFDC) benefits prior to the experiment. Generally speaking, the sample may be characterized as representing what is commonly referred to as the "working poor."

III. EFFECTS OF THE EXPERIMENT ON WORK EFFORT

In estimating the work effort response to SIME/DIME, we have adopted an approach that enables us to distinguish the effects of changing guarantee levels and tax rates. Referring again to Figure 1, it is seen that the experiment increases the disposable income of all families with gross incomes below the tax breakeven level.[4] Economic

[4] Families with gross income below the tax breakeven level are called program participants.

theory predicts that an increase in income that is not work related will induce an individual to reduce the amount of time spent working because leisure becomes more attractive. For purposes of analyzing the effects of the income maintenance experiments, we term the change in work effort associated with an increase in income the *guarantee effect.*

TABLE 2.—SELECTED CHARACTERISTICS OF BLACK AND WHITE FAMILIES AT ENROLLMENT

	Husbands	Wives	Female heads
Average normal income per year	$6,660	$6,660	$3,950
Average hours worked per year	1,719	559	1,010
Average hourly wage rate among workers	$3.30	$2.21	$2.42
Percent employed	80	41	56
Percent previously receiving welfare benefits (AFDC)	14	14	54
Percent in Denver	49	49	49
Average age	34	31	34
Average years of education	11.6	11.5	11.5
Percent black	40	40	56
Average number of family members	4.3	4.3	3.5
Percent control families	47	47	39
Average initial payment per year for families below the breakeven level [1]	$1,330	$1,330	$1,160

[1] This amount excludes AFDC benefits received prior to enrollment that are reimbursed by the experiment.

The experiment also increases the tax rate an individual faces.[5] Again, economic theory predicts that an increase in the tax rate (holding disposable income constant) induces an individual to reduce the amount of time spent working because a higher tax rate implies a lower economic return to working. We term the change in work effort associated with an increase in the tax rate, holding disposable income constant, the *compensated tax effect.*

By adopting an approach that identifies guarantee and tax effects, alternative income support programs can be compared with respect to the two program parameters that are set independently by public policy. Thus, it is possible to estimate the disincentive effects of several competing programs; information that is useful in designing an optimal program.

Table 3 presents estimated compensated tax and guarantee effects on annual hours of work for participants in the Seattle and Denver Income Maintenance Experiments. These estimates apply to heads of families who were employed prior to the experiments and who remain employed during the experiments.[6]

The figures in Table 3 indicate a modest disincentive effect for husbands and a substantial disincentive for wives and female heads of

[5] For persons who received public transfers (such as AFDC and Food Stamps) prior to the experiment, the tax rate may actually be lower under the experiment.

[6] We performed tests to determine whether the estimated responses differed by race, site, and experimental duration. The test results suggested that they do not.

families. Percentagewise, the affects are −5% for husbands, −22% for wives, and −11% for female heads. For men, the total response is about equally divided between guarantee and tax effects; while for women, most of the effect is due to the guarantee. It is important to recognize, however, that these experimental effects are based on guarantee levels and tax rates resulting from the set of programs being tested in SIME/DIME, and not from any single income maintenance program. Furthermore, because the distribution of income in the experimental sample is considerably different from the distribution of income in the U.S. population, the same set of programs tested at the national level may have a substantially different effect.

In addition to causing a reduction in annual hours of work for persons employed, the experiment also reduces the probability of employment. The probability of employment can be reduced either by lengthening the period of time between jobs or by shortening the period of time spent on a given job. Table 4 presents estimates of the effects of the experiment on the probability of employment. For husbands, there is a very small reduction in the probability of employment which stems about equally from shortening periods of time on a given job and lengthening periods of time between jobs. Thus, husbands in the experimental group tend to remain unemployed for slightly longer periods of time and tend to hold jobs for slightly shorter periods of time than husbands in the control group.

Wives and female heads of families exhibit a somewhat larger reduction in the probability of employment than husbands. The reduction for women stems almost entirely from lengthening each period of time spent not employed. An implication of these results is that women in the experimental group who were not employed prior to the experiment were less likely to seek employment during the experiment than women in the control group; while women in the experimental group who were employed prior to the experiment were only slightly more likely to leave employment than women in the control group. We have not yet analyzed how women spent this additional time. Because most women in SIME/DIME have young children, it is likely that a large part of the additional time was spent in productive activities in the home (such as child rearing), rather than in active job search.

TABLE 3.—TAX AND GUARANTEE EFFECTS ON ANNUAL HOURS OF WORK FOR THE AVERAGE WORKING INDIVIDUAL BELOW THE BREAKEVEN LEVEL

	Husbands	Wives	Female heads
Tax effect	−56	−64	−59
Guarantee effect	−47	−199	−117
Total effect	−103	−263	−176
Percentage effect	−5	−22	−11

Source: Michael C. Keeley, Philip K. Robins, Robert G. Spiegelman, Richard W. West, "The Labor Supply Effects and Costs of Alternative Negative Income Tax Programs: Evidence from the Seattle and Denver Income Maintenance Experiments: Part I, The Labor Supply Response Function," Research Memorandum 38, Center for the Study of Welfare Policy, Stanford Research Institute, Menlo Park, Calif., May 1977.

TABLE 4.—EFFECTS OF THE EXPERIMENT ON THE PROBABILITY OF WORKING AND ON THE LENGTH OF TIME SPENT WORKING AND NOT WORKING

	Husbands	Wives	Female heads
Probability of working in the absence of the experiment	0.79	0.40	0.55
Experimental effect	−0.02	−0.07	−0.07
Source of experimental effect:			
Percent change in the length of time spent working	−7	7	−3
Percent change in the length of time spent not working	7	55	48

Source: Philip K. Robins and Nancy Brandon Tuma, "Changes in Rates of Entering and Leaving Employment Under a Negative Income Tax Program: Evidence from the Seattle and Denver Income Maintenance Experiments," Research Memorandum 48, Center for the Study of Welfare Policy, Stanford Research Institute, Menlo Park, Calif., March 1977.

IV. IMPLICATIONS OF THE WORK EFFORT RESULTS FOR A NATIONAL PROGRAM

In order to make use of the information provided by the experiments in the design of a national program, it is necessary to extrapolate the experimental results to the national population. We have used the technique of microsimulation to generalize the experimental results.

Microsimulation consists of applying social program regulations and behavioral assumptions to a data base containing disaggregated information about individuals or groups in order to project program costs and caseloads under varying conditions. To generalize the SIME/DIME results, we use the Micro Analysis of Transfer to Households (MATH) model to assess the effects of a variety of nationwide negative income tax programs. The MATH model reproduces program eligibility requirements and benefit determination schedules. It also estimates behavior of low-income families regarding welfare participation and work effort.

The tax liability, transfer payment, and amount of employment are determined for each family both before and after the NIT is implemented, and the results are summed to derive the total change in costs, caseloads, and work effort under alternative plans. The different effects on various family types are also determined. Six NIT plans with varying tax rates and levels of support are simulated using the March 1975 Current Population Survey (CPS).

The income data from the March 1975 CPS are annual data for the year 1974. Thus, the calculations represent what the effects of an NIT would have been in 1974. No attempt is made to update the responses or cost estimates to later years.

The six programs for which predictions are made have constant tax rates of 50 percent and 70 percent on earnings, and support (guarantee) levels of 50 percent, 75 percent, and 100 percent of the poverty

level ($5,000 for a family of four in 1974). Because the proverty level increases with family size, the support level also increases with family size. The nominal support level is assumed to be constant across regions. The NIT replaces the existing AFDC and Food Stamps programs, taxes all other nonlabor income at the rate of 100 percent, and reimburses positive income taxes below the tax breakeven level. All families that are eligible to receive benefits are assumed to participate.[7]

Work Effort Responses to a Nationwide Program

The average work effort responses to the six nationwide NIT programs are presented in Table 5. The results are reported in two ways: first, the average responses for all participating families, i.e., families receiving benefits from the program; and second, the average responses for the U.S. population. The average responses for the U.S. population include non-responders, as well as responses of participants and nonparticipants. The nonparticipants who respond are families that previously received welfare benefits and are above the breakeven level of the NIT program. These families increase their work effort when the walfare programs are replaced by the NIT program.

In interpreting the results, it is important to keep in mind that the responses vary not only because of changing guarantee levels and tax rates, but also because of a changing pool of participants. For example, as the tax rate increases (for a given guarantee), the pool of participants decreases. The manner in which the pool changes depends on the distribution of income within the revelant population subgroup. For the programs simulated, the number of participating families (e.g., those who receive benefits) ranges from 3.3 million to 19.3 million.

---

[7] The NIT program with a support level equal to 75 percent of the poverty level and a tax rate equal to 50 percent is the program most comparable to the cash assistance portion of the PBJI, with several important exceptions. First, the support level of the PBJI, is only 65 percent of the poverty level. Second, under the PBJI, families receive lower benefits if their earnings are under $3,800 per year and the primary earner is expected to work. (In our simulations we do not impose a work requirement.) Third, the PBJI, taxes most nonlabor income at the rate of 80 percent (Federal assistance is taxed at the rate of 100 percent). Fourth, the tax reimbursement provisions of the PBJI are somewhat less generous than the tax reimbursement procedure used in the simulations (only Federal income taxes are reimbursed under the PBJI). Fifth, the PBJI contains an extension of the Earned Income Tax Credit (EITC) which tends to reduce program tax rates. Sixth, the simulations do not assume that welfare families made worse off by the NIT are given supplemental benefits. Because of these and other differences between the simulated programs and the PBJI, the figures presented in this paper should not be interpreted as representing estimates of the work effort effects and costs of the PBJI. The figures are presented primarily to compare the work effort effects and costs of alternative income maintenance programs.

TABEL 5.—AVERAGE LABOR-SUPPLY RESPONSES FOR ALL PARTICIPATING FAMILIES AND FOR ALL FAMILIES IN THE UNITED STATES

NIT support level	NIT tax rate 50 percent					NIT tax rate 70 percent				
	Participating families			All U.S. families		Participating families			All U.S. families	
	Change in annual hours of work	Percent change	Participating families (millions)	Change in annual hours of work	Percent change	Change in annual hours of work	Percent change	Partitipating families (millions)	Change in annual hours of work	Percent change
50 percent of poverty level:[1]										
Husbands	104	−7.0	.....	−4	−0.2	−136	−10.8	.....	−2	−0.1
Wives	−92	−23.3	.....	−2	−.3	−111	−29.9	.....	0	0
Total (H+W)	−196	−10.3	2.4	−6	−.2	−247	−15.1	1.3	−2	−.1
Female heads	0	0	2.3	+16	+1.6	−10	−2.7	2.0	+20	+2.0
75 percent of poverty level:[1]										
Husbands	−106	−5.9	.....	−19	−1.0	−157	−11.2	.....	−9	−.5
Wives	−110	−22.8	.....	−19	−2.4	−126	−32.5	.....	−5	−.6
Total (H+W)	−216	−9.5	7.6	−38	1.4	−283	−15.8	2.8	−14	−.5

TABLE 5.—AVERAGE LABOR-SUPPLY RESPONSES FOR ALL PARTICIPATING FAMILIES AND FOR ALL FAMILIES IN THE UNITED STATES—Continued

NIT support level	NIT Tax Reate 50 percent					NIT tax rate 70 pefcent				
	Participating families			All U.S. families		Particpating families			All U.S. families	
	Change annual in hours of work	Percent change	Partici-pating families (millions)	Change in annual hours of work	Percent change	Change in annual hours of work	Percent change	Partiti-pating families (millions)	Change in annual hours of work	Percent change
Female heads.....	—47	—6.7	3.0	—23	—2.4	—47	—9.3	2.5	—12	—1.2
100 percent of poverty level: [1]										
Husbands.....	—119	—6.2	.....	—47	—2.4	—164	—10.1	.....	—23	—1.2
Wives.....	—130	—22.7	.....	—50	—6.3	—144	—32.0	.....	—18	—2.3
Total (H+W).....	—249	—10.0	15.7	—97	—3.5	—308	—20.6	5.8	—41	—1.5
Female heads.....	—99	—12.0	3.6	—69	—7.1	—95	—14.9	3.0	—52	—5.3

[1] Poverty level is $5,000 per year for a family of four in 1974.

Note: Average hours of work per year before response, all husbands in the United States = 1,999. Average hours of work per year before response, all wives in the United States = 793. Total number of husband-wife families in the United States = 39,800,000. Average hours of work per year before response, female heads in the United States = 974. Total number of female-headed families in the United States = 4,900,000.

Source: Michael C. Keeley, Philip K. Robins, Richard W. West, "The Labor Supply Effects and Costs of Alternative Negative Income Tax Programs: Evidence from the Seattle and Denver Income Maintenance Experiments: Part II, National Predictions Using the Labor Supply Response Function," Research Memorandum 39, Center for the Study of Welfare Policy, Stanford Research Institute, Menlo Park, Calif. May 1977.

For participating husband-wife families, the magnitudes of the average responses are positively associated with both the guarantee and the tax rate. For participating female-headed families, the responses are positively associated with the guarantee, but do not vary with the tax rate. For both groups, the results indicate fairly sizeable reductions in work effort, ranging from between 10 percent and 21 percent for husband-wife families and between 0 percent and 15 percent for female-headed families.

The average responses of the U.S. population are quite small relative to the average responses of participating families because most families in the United States do not participate in the program. While the magnitude of the average responses increases with the guarantee, it decreases with the tax rate for both groups. This inverse relationship between the average U.S. response and the tax rate is an interesting and perhaps unexpected result that is a consequence of the fact that the number of participants decreases by an amount large enough to offset the effect of a larger response among participants. Thus, the total disincentive effect of a nationwide NIT program is smaller under higher tax rate programs, despite the fact that the response of participating families is larger.

Costs of a Nationwide Program

Estimated annual program costs are presented in Table 6. Program costs are defined to be net of the current costs of the AFDC, AFDC-UP, and Food Stamps programs, which are replaced by the NIT.

The costs of a nationwide NIT vary widely with the parameters of the program. The most expensive program (support level equal to 100 percent of the poverty level and tax rate equal to 50 percent) costs $30 billion more than the current welfare system, and has approximately 39 percent of all husband-wife families and 73 percent of all female-headed families participating in the program. The least expensive program (support level equal to 50 percent of the poverty level and tax rate equal to 70 percent) costs $4 billion less than the current welfare system (which represents a 41 percent savings in welfare program costs) and has approximately 3 percent of all husband-wife families and 41 percent of all female-headed families participating in the program.

TABLE 6.—PROGRAM COSTS BEFORE AND AFTER RESPONSE, HUSBAND-WIFE AND FEMALE-HEADED FAMILIES

NIT support level	NIT tax rate 50 percent				NIT tax rate 70 percent			
	In billions				In billions			
	Program costs before response	Change in program costs due to response	Program costs after response	Participating families (millions)	Program costs before response	Change in program costs due to response	Program costs after response	Participating families (millions)
50 percent poverty level:[1]								
Husband-wife families	—$0.1	$0.3	$0.2	2.4	$—0.8	$0.2	—$0.6	1.3
Female-headed families	—2.9	—.1	—3.0	2.3	—3.3	0	—3.3	2.0
Total	—3.0	.2	—2.8	4.7	—4.1	.2	—3.9	3.3
75 percent of poverty level:[1]								
Husband-wife families	5.4	2.2	7.6	7.6	1.6	1.1	2.7	2.8
Female-headed families	.2	.2	.4	3.0	—.6	.1	—.5	2.5
Total	5.6	2.4	8.0	10.6	1.0	1.2	2.2	5.3

100 percent of poverty level:[1]								
Husband-wife families.....	19.0	6.5	25.5	15.7	6.5	3.1	9.6	3.8
Female-headed families.....	4.0	.5	4.5	3.6	2.6	.4	3.0	3.0
Total.....	23.0	7.0	30.0	19.3	9.1	3.5	12.6	8.8

[1] Poverty level is $5,000 per year for a family of 4 in 1974.

Note: Total number of husband-wife families in the United States equals 39,800,000. Total number of female-headed families in the United States equals 4,900,000.

Source: Michael C. Keeley, Philip K. Robins, Richard W. West "The Labor Supply Effects and Costs of Alternative Negative Income Tax Programs: Evidence from the Seattle and Denver Income Maintenance Experiments: Part II, National Predictions Using the Labor Supply Response Function," Research Memorandum 39, Center for the Study of Welfare Policy, Stanford Research Institute, Menlo Park, Calif. May 1977.

For programs with positive costs, the proportion due to the work effort response varies between 23% and 55%. The magnitude of these additional costs demonstrates the importance of accounting for work effort adjustments when designing a national program. Failure to take work effort adjustments into account can lead to a serious underestimate of total program costs.

Effects on the Welfare Population

Since the simulations assume that certain welfare programs (AFDC, AFDC–UP, Food Stamps) are replaced by the NIT, and that there is no state supplementation of lost welfare benefits, it is likely that some families are made worse off by the program (i.e., their disposable income is reduced). Table 7 presents a tabulation of the number and percentage of welfare families that are made worse off by the NIT, assuming no state supplementation. As this table indicates, the percentages are quite large, even for the more generous NIT programs. For example, under an NIT program with a support level equal to the poverty level and a tax rate equal to 50%, one quarter of the welfare families are made worse off. To compensate families made worse off by the NIT would likely result in a substantial increase in program costs.

The reason why so many families are made worse off by the NIT may be due to the fact that there are loopholes in the existing welfare system that enable families to face very low benefit reduction rates.[8] These low benefit reduction rates imply that welfare grants remain high even when family members work a substantial number of hours. Thus, even though the support level of the NIT may be higher than the support level of welfare, the higher NIT tax rates makes many working welfare families worse off.

[8] The main loophole arises from generous provisions regarding the deduction of work related expenses from income.

TABLE 7.—NUMBER AND PERCENTAGE OF WELFARE [1] FAMILIES MADE WORSE OFF BY THE NIT NO STATE SUPPLEMENTATION

NIT support level	NIT tax rate 50 percent		NIT tax rate 70 percent	
	Number made worse off (millions)	Percent made worse off	Numb r made worse off (millions)	Percent made worse off
50 percent of poverty level [2]				
Husband-wife families..	1.2	79	1.4	89
Female-headed families.	1.8	93	1.9	95
Total.....	3.0	87	3.3	92
75 percent of poverty level: [2]				
Husband-wife families..	.7	43	1.2	71
Female-headed families.	1.4	67	1.6	75
Total.....	2.1	59	2.8	73
100 percent of poverty level: [2]				
Husband-wife families..	.4	23	.7	41
Female-headed families.	.5	25	.7	33
Total.....	.9	24	1.4	37

[1] AFDC, AFDC-UP, food stamps.
[2] Poverty level is $5,000 per year for a family of 4 in 1974.

Source: Simulation runs prepared for SRI International by Mathematica Policy Research and the Hendrickson Corp.

V. EFFECTS OF THE EXPERIMENT ON MARITAL STATUS

SIME/DIME has provided an opportunity to study the effects of an NIT on marital dissolution. We begin by reviewing the reasons for suspecting that an NIT will affect rates of marital disruption.

First, an NIT would remove the incentives to marital dissolution inherent in the current system. Under certain circumstances, the income available to a family can increase if the husband is not present in the home. There have been no empirical studies that demonstrate that these incentives have any effect upon dissolution rates. However, such incentives would not be present in an NIT program in which eligibility and benefit levels are not dependent upon family composition. Thus, in any effect these incentives would not be present in an NIT.

A second reason for expecting an NIT to alter dissolution rates rests on the observed association between family income and marital dissolution rates. Many studies have shown that the probability of marital dissolution is highest for the lowest income families. If poor families have high rates of marital dissolution not because they lack material resources, but because they lack appropriate values and personality traits, then altering income levels will not greatly affect marital stability in this population.

On the other hand, many argue that income levels affect the ability of the families to cope with a variety of problems and dissatisfactions. Further, it is argued that male heads of families who cannot provide certain consumption standards for their families are viewed as failures by themselves and others. One response to such failure is flight from marriage relationship. Income supplement programs that substantially improve living standards might reduce the pressure towards dissolution. We refer to effects of this sort as income effects. We expect that the income effects of an NIT would lower the rate of marital dissolution.

But there is another effect of an NIT that has been overlooked in most policy discussion. Early in our research we suggested that an NIT would alter the structure of dependence in marriages (Hannan, Beaver, and Tuma, 1974). An NIT guarantees support to unmarried as well as married. As a result, an NIT will alter the level of resources available outside of marriage and thereby alter the dependence of the members on marriage. We refer to this effect as the independence effect. Since the NIT increases the level of resources outside of marriage, the independence effect will raise the probability of marital dissolution.

A final issue to consider is welfare discounting. If participation in the current system is degrading, both its income and independence effects are muted. Families receiving payments would not experience the full income effect due to the strain induced by stigma. Likewise, dependent spouses would not experience the full independence effect of the welfare system if it is viewed as degrading. This suggests that a payment from an NIT program will have a stronger income and independence effect than a payment of the same amount from welfare. Another way of putting this is to say that welfare is "discounted" in its effects on marriage relative to an NIT.

There are other nonpecuniary differences between welfare and NIT programs that may result in welfare being discounted. Participation in the NIT involves less effort than going on welfare. Our experimental NIT program has a simpler and presumably less alienating bureaucracy. The rules of the NIT are carefully explained to the participants. Information about eligibility rules and support levels of

welfare may not be as well known. Any of these three factors (stigma, transaction costs, or lack of information) suggest that the effects of welfare may be discounted.

What, then can be said about the expected impact of an NIT on marital dissolution rates? For an NIT that is more generous than the present welfare system, as is the case with the Seattle-Denver experiment, it is not possible to predict the direction of the impact a priori. If the income effects dominate, the NIT will lower the dissolution rate. If the independence effects are stronger the reverse will hold. Even a less generous program may have both income and independence effects if the changes in the program affect the rate at which welfare is discounted.

Basic Experimental Findings

Our findings show that the NIT program destabilizes marriage. Controlling for the variables used in assigning families to treatments and several other variables that may affect dissolution, we found that the experiment significantly increases the dissolution rate for both Whites and Blacks. The differences between the experimental and the control groups are statistically significant for both races indicating that we can with some confidence rule out the possibility that the experimental-control difference is due merely to chance.

This finding is consistent with a model in which the independence effects dominate the income effects for the programs tested. But does it imply that all NIT schemes will increase dissolution rates in populations like those we studied? To answer this question we must consider some more complex analyses. Our most provocative findings concern the patterns of impacts by level of income support. The lowest support level holds particular interest since it differs little in financial terms from the existing level of support available from the AFDC and Food Stamps. If welfare is not discounted, this program should have no independence effect. But the dissolution rate for families on this treatment greatly exceeds that of the control groups—by 96 percent for Whites, by 67 percent for Blacks, and by 60 percent for Chicanos (see Table 8).* So we conclude that the independence effects of welfare are indeed discounted relative to those of an NIT. A curious results shown in Table 8 is that, for each race-ethnic group, the plan that guarantees income at the highest level, 140 percent of the poverty line, has the smallest impact. These findings make plain the need to understand the stigma and information content of NIT schemes in order to compare their effects with the existing system.

The basic results of the experimental analysis are robust. We found no technical problem that explains away the findings. One problem deserves mention: attrition. We lost track of some families, and others refused to participate after a time. We suspected that a family's decision to remain in the study was affected both by the benefits they receive from the experiment and by marital events. If control families were more likely to leave the experiment at the time of a marital dissolution, our records would undercount dissolutions for this group. This bias would inflate experimental-control differences. Luckily, the attrition

* Throughout this paper, we report impacts estimated over the first 2 years of the experiment for 5-year experimental families. The effects for 3-year families are approximately 80 percent of the 5-year effects.

rates in this experiment are low, about 10 percent over two years. But in studies of rare events such as marital disruptions, even small attrition rates may give misleading results. So we investigated the sensitivity of our results to attrition (Hannan, Tuma and Groeneveld, 1976). They are not very sensitive. Even if all the controls who left the experiment had an unrecorded marital dissolution, the experimental-control difference would still be positive and significant for Whites and Blacks. The difference between the low support treatment group and the controls is the most robust of all the basic findings.

TABLE 8.—PERCENT CHANGE IN MARITAL DISSOLUTION RATE BY LEVEL OF INCOME GUARANTEE

Guarantee level	Race-ethnic group		
	Black	White	Chicano
90 percent of poverty line	[1] 67	[2] 96	60
125 percent of poverty line	[2] 93	[1] 55	−28
140 percent of poverty line	21	12	−.35
Average of SIME/DIME program	[2] 61	[1] 58	−4
Number of cases	939	1,297	518

[1] Significant at the 0.05 level.
[2] Significant at the 0.01 level.

The Income and Independence Effects of an NIT

To probe the mechanism creating these experimental effects, we attempted to parameterize the income and independence effects. Recall that the 90 percent of poverty level support has a larger impact on dissolution than that of the 140 percent level. Moreover, the former is statistically significant while the latter is not. Why does a small financial change from the control environment have a strong impact when a bigger change does not?

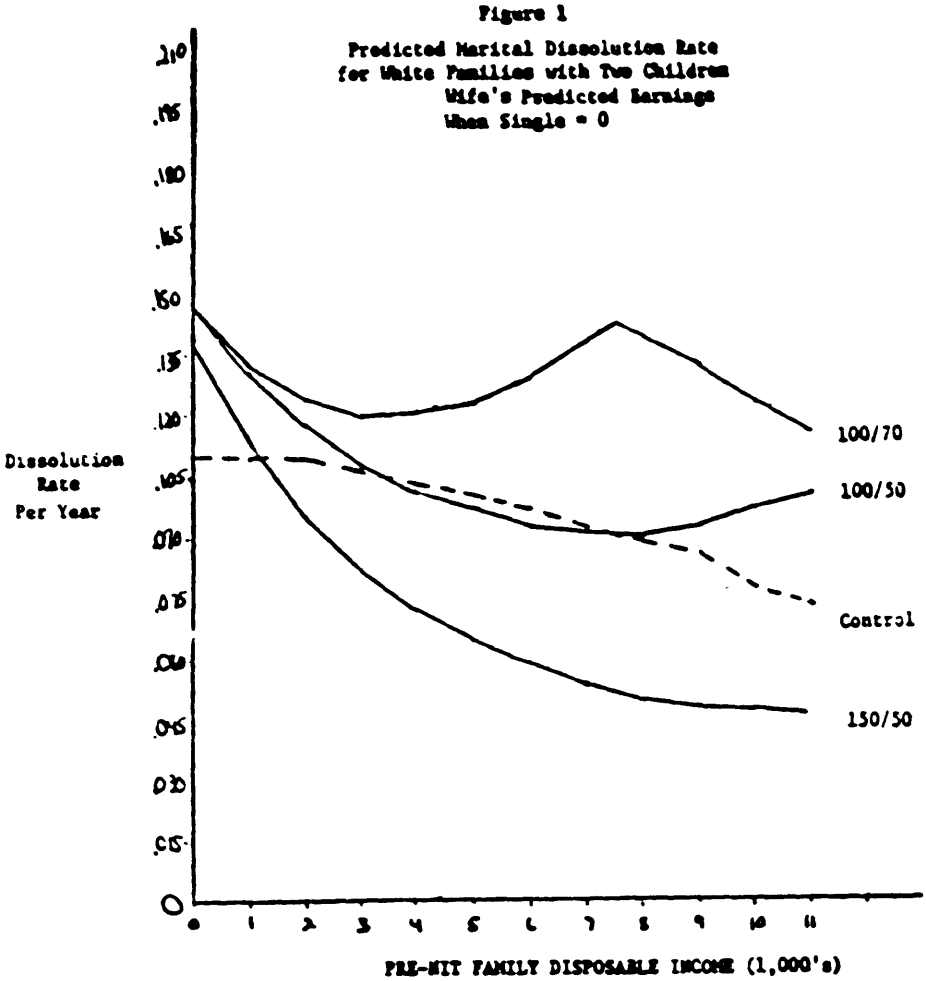
We sought to explain this pattern of experimental-control difference with a model of the income and independence effects of the NIT program. Our model and the evidence supporting it are discussed at length elsewhere (see Hannan, Tuma, and Groeneveld [1977a, 1977b]). Briefly, our model assumes that the income and independence effects are nonlinear functions of income.

We address the problem by using our model for the income and independence effects of NIT payments. Our analysis reveals that the impact of any NIT program differs according to the race-ethnicity of the family, the number of children, and a variety of other demographic and background characteristics. So we must calculate impacts separately for each combination of characteristics. We cannot be exhaus-

tive here but will illustrate the impacts of various NIT programs on rates of marital dissolution for white families with two children in which each spouse is aged 25, has 12 years of education, and the couple has been married for 5 years. We will vary both family income prior to the NIT and wife's pre-NIT independence. As will become clear, the latter plays a crucial role in determining the NIT impact on dissolution rates. We consider two cases typical of those we studied: (1) wives who would not be employed upon becoming single; (2) wives who would earn $3000 per year as a single woman. In each case we assume, in line with the discussion earlier and our empirical findings, that welfare is "discounted". In particular, we assume that each dollar of welfare guarantee has an independence effect half as large as that of a dollar of earnings or a dollar from the NIT program.[10]

Figure 1

Predicted Marital Dissolution Rate
for White Families with Two Children
Wife's Predicted Earnings
When Single = 0



Figures 1 and 2 plot the dissolution rate under various programs by levels of pre-NIT family disposable income. Figure 1 contains the predicted curve for families in which the wives would have no earnings after leaving the marriage. Consider the most generous NIT program depicted in Figure 1, the 150 percent of poverty level support with 50 percent tax, denoted 150/50. It is below the control curve almost

[10] We have tried other discounts and found that as long as welfare is discounted, the experimental impacts can be explained. However, the 50 percent discount gives better results than others we tried.

everywhere. For most families, the stabilizing effect of the program outweighs the independence effect. Only for the poorest families is this not so. For them the income effect curve is quite flat. Even large changes in family income have relatively small stabilizing effects; consequently, the independence effect dominates in the low range of family incomes.

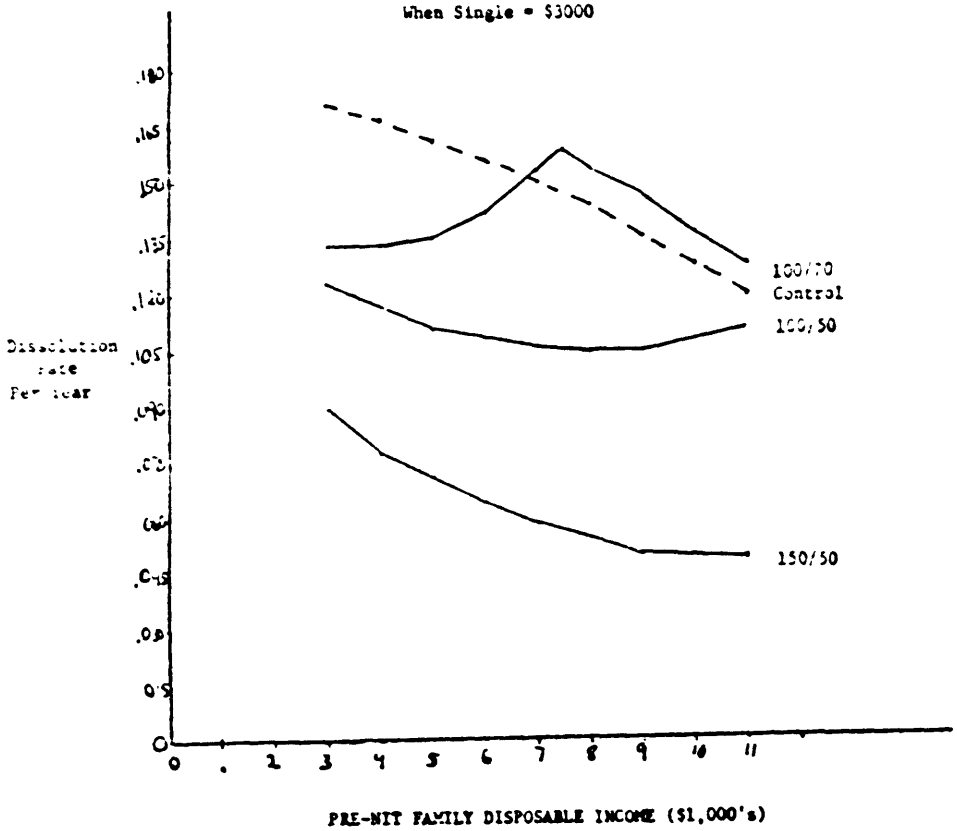
Next, examine the 100 percent of poverty level, 70 percent tax program, denoted 100/70. The curve for this plan is above the control curve at all levels of pre-NIT family income. In other words, the plan does not raise family income sufficiently to induce income effects strong enough to offset the independence effects of the program.

Finally, consider the 100 percent of poverty level, 50 percent tax program (100/50). It has a dissolution curve that falls between those of the other two NIT programs. It gives higher dissolution rates than a program with the same tax rate but a higher support level. This outcome reflects the curvature of the income and independence effects. Both effects are increased when one increases the support level from 100 percent to 150 percent. However, the income effect of such an increase dominates the increase in the independence effect.

The 100 percent of poverty level, 50 percent program (100/50) gives lower dissolution rates than a program with the same support level but a high tax rate. Increasing the tax rate reduces the income effect of

Figure 2

Predicted Marital Dissolution Rate
for White Families with Two Children
Wife's Predicted Earnings
When Single = $3000



the program because families receive smaller payments at any level of pre-NIT family income. It does not, however, lessen the independence

effect for women who have no earnings. Thus, according to our model the 100/70 program has the same independence effect as the 100/50 program, but it has a smaller income effect.

Above we mentioned that wife's independence before the NIT is important in determining the NIT impact. We see this in Figure 2, which plots curves for the same programs as in Figure 1, but for families in which the wife would earn $3000 per year after leaving the marriage. Now the various NIT programs mainly decrease the dissolution rate; that is, the increase in independence that they induce is small relative to their effects on improved family well-being. Curves for both the 150/50 and 100/50 plans are below the control curve for the range of pre-NIT family incomes plotted. The NIT increases the rate of dissolution only for the 100/70 plan, and then only for families with pre-NIT income above $6,500. Furthermore, the increase due to the NIT is rather modest.

Several general tendencies emerge from these and other figures not reported here. First, the NIT impact is mainly concentrated in those families with the most dependent wives. For working women, introduction of an NIT changes only slightly the quality of financial alternatives to an existing marriage, and thereby has less impact on decisions to end a marriage. Second, the high support and low tax programs yield the lowest dissolution rates, and, at least for Whites, these are normally below the pre-NIT rates. Plans with lower support levels and higher tax rates tend either to be closer to the control curve or to increase the dissolution rate.

VI. CONCLUSIONS

The following summarizes the major conclusions of this paper.

A. Within the SIME/DIME sample, there is a modest decline in work effort among male heads of families (5 percent) and a substantial decline among spouses (22 percent) and single female heads (11 percent).

B. A large proportion of the reduction in work effort among women represents time out of the labor force.

C. The work effort response to a nationwide NIT program is very sensitive to the program support level and tax rate.

D. Failure to take work effort response into account when designing a national program can lead to a serious underestimate of total program costs.

E. The total work effort response to a nationwide NIT program is smaller under higher tax rate programs, despite the fact that the work effort response among participating families is larger.

F. The total costs of a nationwide NIT program are very sensitive to the support level and the tax rate.

G. Compensation of welfare families made worse off by a nationwide NIT program is likely to result in a substantial increase in program costs.

H. The NIT plans tested in SIME/DIME tend to substantially increase the rate of marital dissolution among Black and White families.

I. The greatest increase in marital dissolution occurred at the lowest support levels and the smallest increase occurred at the highest support level.

J. The experimental impact on marital dissolution appears to be operating through offsetting income and independence effects. The dominance of the independence effect at low support levels, plus the tendency for married women to partially discount the potential benefits from the existing welfare system, could explain the high impact of the low NIT support levels.

References

Hannan, Michael T., Steven E. Beaver, and Nancy B. Tuma. "Income Maintenance Effects on the Making and Breaking of Marriages: Preliminary Analysis of the First Eighteen Months of the Seattle and Denver Income Maintenance Experiment." Mimeographed. Center for the Study of Welfare Policy, Stanford Research Institute, Menlo Park, California, 1974.

Hannan, Michael T., Nancy Brandon Tuma, and Lyle P. Groeneveld. "The Impact of Income Maintenance on the Making and Breaking of Marital Unions: An Interim Report". Research Memorandum 28, Center for the Study of Welfare Policy, Stanford Research Institute, Menlo Park, California, June 1976.

Hannan, Michael T., Nancy Brandon Tuma, and Lyle P. Groeneveld. "Income and Marital Events: Evidence from an Income Maintenance Experiment." American Journal of Sociology 82, May 1977a: 1186–1211.

Hannan, Michael T., Nancy Brandon Tuma, and Lyle P. Groeneveld. "A Model for the Effect of Income Maintenance on Rates of Marital Dissolution: Evidence from the Seattle and Denver Income Maintenance Experiments." Research Memorandum 44. Center from the Study of Welfare Policy, Stanford Research Institute, Menlo Park, California, February 1977b.

Keeley, Michael C., Philip K. Robins, Robert G. Spiegelman, Richard W. West, "The Labor Supply Effects and Costs of Alternative Negative Income Tax Programs: Evidence from the Seattle and Denver Income Maintenance Experiments: Part I. The Labor Supply Response Function". Research Memoradum 38. Center for the Study of Welfare Policy. Stanford Research Institute, May 1977a: forthcoming in American Economic Review, December 1978, under the title "The Estimation of Labor Supply Models Using Experimental Data".

Keeley, Michael C., Philip K. Robins, Robert G. Spiegelman, Richard W. West, "The Labor Supply Effects and Costs of Alternative Negative Income Tax Programs: Evidence from the Seattle and Denver Income Maintenance Experiments: Part II. National Predictions Using the Labor Supply Response Function". Research Memorandum 39. Center for the Study of Welfare Policy, Stanford Research Institute, May 1977b; also in Journal of Human Resources, Volume XIII, Winter 1978, pp. 3–36.

Robins, Philip K. and Nancy Brandon Tuma. "Changes in Rates of Entering and Leaving Employment Under a Negative Income Tax Program: Evidence from the Seattle and Denver Income Maintenance Experiments". Research Memorandum 48. Center for the Study of Welfare Policy, Stanford Research Institute, March 1977.

V. Statement of Dr. Robert G. Williams, Vice President, Mathematica Policy Research, Inc.*

INTRODUCTION

Mr. Chairman and Members of the Committee, I appreciate this opportunity to appear before you today to discuss the work my firm has been doing to assist the State of Colorado and the Department of Health, Education and Welfare in improving the administration of public welfare programs. We believe that this work has important implications, not only for the operation of current income transfer programs such as AFDC and Food Stamps, but also for the design and implementation of future programs that might be developed as a result of current welfare reform efforts.

For the past two years, we have been involved in the development and operation of the Colorado Monthly Reporting Experiment. This project is a major test of an alternative administrative system intended to improve the accuracy and responsiveness of payments to recipients in the AFDC program.

Preliminary research results from the project indicate that it has been remarkably effective in attaining these goals. There is strong evidence that this alternative administrative system, based on a monthly retrospective reporting principle, has increased the responsiveness of the program to recipients by identifying and reacting to many changes in recipient circumstances that are missed under the traditional administrative system. One finding indicates that in a given month the monthly reporting system processes grant changes for two and one-half times as many cases as does the traditional AFDC system. This implies strongly that the traditional AFDC system misses many changes in recipient circumstances and is unresponsive to their changing needs. There is substantial evidence that the monthly reporting system is more accurate as well. A direct comparison of payments to cases in this new system and payments to a comparable control group in the traditional system shows that monthly reporting reduces payments by about 8 percent. This result is attained without any tightening of eligibility standards or reduction of benefits to eligible families. The monthly reporting system apparently identifies ineligible recipients as soon as their status has changed, and curtails payments immediately rather than after the lags of as much as several months which often occur under the traditional system. These goals have been attained with only a negligible increase in administrative costs, which are greatly overshadowed by savings in benefits.

These preliminary results offer considerable encouragement that substantial improvements can be made in the administration of our in-

[1] Opinions expressed are those of the author and do not necessarily represent the view of sponsoring agencies.

*Statement at hearings before Subcommittee on Government Operations, House of Representatives, 1977.

come transfer programs serving needy families. In the remainder of my statement. I will be describing the nature of the Colorado project in more detail, reviewing the specific impact of the monthly reporting system on the responsiveness and accuracy of benefit payments, describing the impact of the system on recipients, and summarizing the effects of the system on administrative patterns and costs.

DESCRIPTION OF COLORADO MONTHLY REPORTING SYSTEM

To present an understandable description of the monthly reporting system being tested in Colorado, I would like to contrast its features with those of the administrative system traditionally used in most areas of the country in determining eligibility and computing benefits for the AFDC program. Although there are differences in detail in the manner that States and counties administer AFDC, there is enough similarity in broad principles that it is accurate enough to refer to a "traditional AFDC system." Comparison of the monthly reporting system in two Colorado counties with the traditional AFDC system used in the rest of Colorado is the source of information on monthly reporting's impact on benefits, recipients, and administrative patterns and costs.

The monthly reporting administrative system developed in Colorado consists of three basic elements: 1) a monthly reporting requirement; 2) a monthly retrospective accounting period; and 3) an automated support system. Under this experiment, recipients must mail in a simple but comprehensive form each month as a requirement for continued eligibility. On this form, called a Monthly Status Report (MSR), recipients are required to report their income, household composition, and other relevant eligibility factors such as school attendance status of children over the age of 16. This monthly reporting requirement contrasts with the much less frequent formal reporting schedule used in the traditional AFDC system. The traditional system requires the completion of forms only every six months for recipients in the regular segment of the AFDC program and only every three months for the smaller number of recipients in the unemployed parent segment of AFDC. Under the traditional administrative system, recipients are, of course, instructed to report changes in circumstances that take place within the intervals between formal reports. However, for a variety of reasons, this informal reporting requirement is often difficult for recipients to interpret and the agencies to enforce. Morever, it is often difficult for recipients to penetrate institutional barriers to report a change and to be sure that agency staff will take the proper action upon any such report.

The monthly retrospective accounting principle can be simply described as computation of each month's grant based on actual circumstances of the recipient in the month prior to payment. In the case of the Colorado Monthly Reporting Experiment, recipients in the calendar month reporting cycle file a Monthly Status Report detailing their actual circumstances for a given month by the 5th of the following month. The data on the Report, as verified and confirmed by the agency, serve as the basis for their next grant payment. In the traditional AFDC administrative system, however, grants are based on the

agency's estimate of a recipient's current need, which is referred to as a prospective accounting principle. Thus, recipients' payments are based on their needs for the month in which the payment is made. Administrative processing lags require that this payment actually be estimated well in advance. Moreover, the length of time between formal reports means that the agency must estimate individual recipient needs for more than six months in advance for most AFDC cases. These estimates, which serve as the basis for payment during that period unless subsequently altered by a recipient-initiated report of change, are computed according to an elaborate set of rules for projecting, averaging, and predicting.

In the Colorado Monthly Reporting Experiment, the Monthly Status Reports filed by recipients provide the input for an automated support system for the project. This support system performs many clerical and bookkeeping functions which traditionally consumed the time of eligibility workers. Upon receipt by the agency, data from the Monthly Status Reports are entered directly into a computer processing system. The system begins by editing the Monthly Status Reports for completeness and consistency and issuing reports of edit problems to eligibility workers. The system also redetermines eligibility; computes grants; produces reduction and discontinuance notices to recipients; generates checks, check registers, and Grant Explanations to recipients; produces case status reports for Technicians and their supervisors; and provides a management summary report of caseload status at the end of the processing cycle. This level of automation contrasts sharply with the level attained in the traditional AFDC system. Although the States and counties vary in the level of automation of administrative functions, the traditional system is characterized by a heavy reliance on manual processing for such functions as eligibility redetermination. grant computation, and transfers of information from form to form.

Perhaps the most important point to be made in describing this difference is that the traditional AFDC systems cannot reach the levels of automation possible under the monthly reporting system because of inherent limitations caused by traditional use of a prospective accounting period and less frequent reporting periods. Since under the traditional system eligibility determination and grant computation require projecting future needs based on a combination of actual past data and recipients' estimates of future needs, human discretion is required to carry out these functions. The data are too imprecise and the rules too ambiguous to permit eligibility determination and grant computation from raw data supplied by the recipient. In the monthly reporting system, however, actual data provided on the Monthly Status Reports is used directly by the computer to make these determinations.

EFFECTS OF THE COLORADO MONTHLY REPORTING SYSTEM ON BENEFITS

The Colorado Monthly Reporting Project is unique among tests of new administrative features in public welfare programs in that it is a true experiment: it provides for a systematic, statistically valid determination of the results of implementing the monthly reporting system. In Denver County, which is typical in most respects of many other urban areas in such terms as socioeconomic composition of the caseload and administrative conditions, ten percent of the caseload (about 1200

cases) has been randomly assigned to the monthly reporting system. Reporting patterns of the monthly reporting group and payments to its members are being compared in a rigorous way to the reporting patterns and payments of a control group in the traditional AFDC administrative system. This control group, like the monthly reporting group, consists of an additional ten percent of the Denver County caseload which is also selected randomly.

Preliminary results from the first eight months of operation in Denver County indicate that the higher frequency of reporting under the monthly reporting system, along with the principle of basing grants on actual circumstances, has resulted in substantial reduction in benefits. Cases in the monthly reporting system received about eight percent less in payments than did cases in the traditional system at the end of this period. This reduction came about without any tightening of eligibility standards and apparently without denying payment to any eligible household. Rather, the reduction in payments seems to have come about primarily because of a reduction in the caseload of a corresponding eight percent. Data from the project suggest that this reduction results from the ability of the monthly reporting system to identify families as soon as changes in their circumstances make them ineligible. Payments to ineligible households are therefore curtailed immediately under the monthly reporting system, whereas frequently a lag in stopping payments occurs in the traditional system.

One of the more surprising and significant findings from this project is that the monthly reporting system has processed several times more grant changes for its caseload than the traditional system for control cases. Even though the two groups of recipients are statistically identical, operation of the monthly reporting system has resulted in computation of grant changes for approximately twenty percent of ongoing cases each month, whereas operation of the traditional system has resulted in computation of grant changes for only about eight percent of the caseload per month. The difference in frequency of grant changes applies equally to both increases and decreases in grants. This finding suggests that under the traditional AFDC system many changes in recipient circumstances are simply missed. Apparently at any given time under the traditional system, an uncomfortably large proportion of AFDC recipients are receiving insufficient assistance to meet their needs and are therefore underpaid, while another large proportion are receiving too much assistance for their needs and are therefore overpaid. The magnitude of these differences is larger than we would have expected, and may help to explain why the AFDC Quality Control process consistently finds unacceptable levels of error in the program. Moreover, even Quality Control findings do not fully document the level of inaccuracy in the traditional system. Quality Control rules tend to exclude from error statistics many mispayments if prescribed recipient reporting periods and administrative action periods have been complied with. Increased frequency of grant changes under the Monthly Reporting Experiment imply strongly that wider implementation of the system being tested in Colorado might substantially improve the level of accuracy and responsiveness of payments in the AFDC program.

IMPACT OF THE COLORADO MONTHLY REPORTING SYSTEM ON RECIPIENTS

There has been widespread concern over the potential impact of this project on recipients. The first concern is that recipients might be unable to meet the filing requirements imposed upon them and that many legitimately in need of assistance would thereby be forced off the program. A second concern is that the retrospective nature of the system, in which grants are based on actual past circumstances, might prove to be unresponsive to recipients suffering sudden setbacks, especially applicants for assistance who frequently have emergency needs. In the Colorado Monthly Reporting Project, careful attention has been given to these concerns from the beginning. Results from operations thus far support the conclusion that a properly designed monthly reporting system does not adversely affect eligible recipients of assistance and that, on balance, the system seems to be more responsive to recipient needs than does the traditional system with which it is being compared.

In the monthly reporting system, recipients must file a mail-in, postage-paid Monthly Status Report each month in order to retain eligibility and provide data for computation of the next payment. The Monthly Status Reports are mailed to recipients three working days before the end of the monthly reporting period. If MSR's are received by the 5th of the following month, the recipient receives payment on the first possible payment date, which is the 16th. Subsequent filing deadlines of the 12th and 20th correspond to second and third payment dates of the 23rd and 30th, respectively. The forms have been designed for comprehensiveness and ease of completion. Recipients must only circle "yes" or "no" responses to individually specified questions, and fill in amounts of income received. No arithmetical computation is required of recipients; they submit only the raw data required for automated calculation. However, recipients must send in paystubs as verification of earnings and provide suitable documentation for other income and changes in household composition.

Experience with the Colorado Monthly Reporting System clearly demonstrates the ability of AFDC families to complete the required monthly reporting forms at an acceptable level of proficiency and to submit them promptly. Generally, more than ninety percent of families who file do so by the first filing deadline, which falls on the 5th of the month—about eight or nine days after the forms are mailed to the recipients by the agency. Most of the remaining families who file do so by the second deadline of the 12th, with only one or two percent submitting forms only in time for the third deadline of the 20th. The level of accuracy on the completed forms is high, with the majority of them suitable to serve as the basis for eligibility redetermination and grant computation directly with no intervention by agency staff.

Recipients demonstrate a definite learning curve in form proficiency. Agency staff noted a certain rate of error in completion of the forms during the first two months of the project. By the third month, however, the level of proficiency improved substantially. The ability of recipients to cope with the forms is undoubtedly aided by their monthly experience in completing them. Because recipients must fill out report forms more frequently than in the traditional system, they

become more familiar with them and find it easier to comply with the filing requirements.

Fewer than five percent of families normally fail to return a Monthly Status Report by the final filing deadline and are therefore discontinued. These families appear to place themselves in this category deliberately since they ignore three separate warnings of the consequences for non-filing. The first warning is prominently displayed on the Monthly Status Report. Second and third warnings, which are formal notices of discontinuance, are sent after non-receipt of the Monthly Status Report on the 5th and the 12th, the first two filing deadlines. Although the agencies involved in this project have been very sensitive to the possibility of adverse effects on recipients, no evidence has been found that any otherwise eligible recipient has been forced off AFDC by the filing requirements of the monthly reporting system. The rate of re-application for assistance has been low and the number of hearings requested has been about normal for the proportion of the caseload involved.

The retrospective payments feature of this monthly reporting system has attracted much scrutiny because of the fear that it could leave recipients short of critical resources when they were faced with an abrupt decline in income. Failure of the monthly reporting system to meet "current need" can be legitimately considered as a potential disadvantage. To minimize this problem, considerable effort was devoted during the design phase to maximizing the processing speed of the system, thereby minimizing the lag between the occurrence of changes in circumstances and receipt of a payment reflecting those changes. The rapid processing schedule has been successful: more than eighty percent of recipients consistently receive payment on the early payment date, the 16th of each month, which is only a half month after the end of the reporting period. There appear to be very few recipients who are adversely affected by the retrospective aspect of the system. On those limited occasions when recipients are caught short by the system, Food stamps are made available with no purchase requirement to supplement other sources of income. New cases receive special treatment because of the frequency of emergency needs. Initial grants are calculated on a current need basis if the retrospective grant determination falls short.

Many persons exaggerate the problems created for recipients by a retrospective system since they compare the actual workings of the monthly reporting system to an ideal rarely approached by the traditional system. Although the traditional system is theoretically intended to meet current need at all times, in practice this ideal is severely compromised by long delays in processing grants. It is compromised further by vagaries of the system that result from the lack of clear procedures, reporting requirements that leave too much agency discretion, and the absence of safeguards to insure that information informally supplied by the recipient will be properly processed. It is our judgment, based on our experience with the monthly reporting system in Colorado, that any disadvantages for the recipient caused by the retrospective principle for computation of grants are more than compensated by the superior responsiveness of the system to changing circumstances and the certainty that recipient reports of such changes will be reflected in adjustments to grants.

IMPACT ON ADMINISTRATIVE FUNCTIONING AND COSTS

Many welfare administrators have been attracted to the concept of monthly retrospective reporting because of its obvious potential for increased accuracy of payments, but have been deterred by the fear that virtually a six-fold increase in written reports by recipients would result in nearly proportionate increases in administrative costs. To test this prospect, and to assess other effects of the monthly reporting system on administrative functioning, a recond phase of the Colorado Monthly Reporting Experiment is a county-wide implementation of the system in Boulder County. Preliminary data from this phase indicate that any increases in administrative costs caused by operating the monthly reporting system are likely to be negligible and that there are many desirable administrative effects of the monthly reporting system.

As I have noted earlier, a characteristic of the traditional AFDC system is a heavy reliance on manual processing. In the usual mode of program administration, a large amount of eligibility staff time is spent performing clerical duties such as computation of grants and filling out forms, often copying substantial amounts of information from one form to another. In contrast, the monthly reporting system developed for Colorado takes full advantage of the potential for automation that is inherent in a monthly retrospective system. Recipient data are entered directly into the computer for machine editing. Once the MSR information is complete and consistent, the computer performs the grant calculation and produces the necessary forms, notices, grant explanations, and checks (among other functions). Because so much of the clerical workload is automated under this system, the increased information flow from recipients can be handled by, at most, the same level of eligibility staff as in the traditional system. It appears that there is some increase required in data entry staff, computer time, postage, and printing costs, although this increase is partly offset by a decrease in the amount of clerical support time required. Our best estimate of the ultimate impact of the system on administrative costs, then, based on several months of operation in Boulder County, is a net increase in the range of zero to ten percent. We believe that these figures may even be upper range estimates of the impact on administrative costs elsewhere since Colorado's level of administrative costs per AFDC case are well below the national average. Thus there might be more available resources to support the system in other jurisdictions and the net impact on administrative costs might be even less. Moreover, there exists a clear potential to expand the monthly reporting system to accommodate a joint administration of AFDC and Food Stamps through the use of a single reporting form and processing system. Exploitation of this potential would most likely bring about a reduction in overall administrative costs through the reduction of duplicate transactions and automation of Food Stamp administrative functions.

Eligibility workers participating in the experiment have noted a number of administrative advantages gained from the monthly reporting system. One of the most significant advantages is increased constructive contact between workers and recipients. The high level of automation in the system has freed workers from many routine clerical

tasks and focused their efforts on obtaining information from recipients and processing the monthly report forms. Workers administering the monthly reporting system in Denver County estimate that their contact with recipients has increased by as much as fifty percent relative to the traditional system. This has permitted the workers to become more aware of recipient circumstances, further improving the accuracy of payments as well as increasing the likelihood that recipients will be referred to need social services.

The monthly reporting system has also had the beneficial effect of clarifying the respective roles of eligibility workers and recipients. For the first time, recipients have clear-cut, unambiguous requirements with which they must comply, as well as a reliable channel for reporting changes in circumstances. (In the monthly reporting system, recipients can even verify that information has been processed correctly They receive computer-printed grant explanations with each check, a source of information that is almost non-existent in the traditional system.) Similarly, workers have the clear obligation to process the reported information, obtain supplemental data from the recipient if necessary, and ensure timely issuance of the grant.

Eligibility workers express a higher degree of job satisfaction under the monthly reporting system. They find their jobs more demanding because of the absolute nature of deadlines and the need to learn certain new skills. However, this very challenge contributes to their feeling of greater satisfaction, as does the better organization of their daily responsibilities, and, most of all, the greater sense of control over their caseloads.

CONCLUSION

The results of the Colorado Monthly Reporting Experiment, while not yet complete, offer considerable encouragement that implementation of a monthly retrospective reporting system similar to that tested in Colorado would represent a major improvement in the administration of the AFDC program. Preliminary estimates from the project indicate that grants to families would decrease about eight percent due to the greater accuracy of the system, and that any increase in administrative costs would be no greater than ten percent. It is important to realize in considering these results that it takes a twelve percent increase in administrative costs in Colorado to offset a one percent decrease in benefits, so that any additional costs of operating a monthly retrospective reporting system would be offset by only a fraction of the reduction in grants. The remaining reduction in grants would thus represent a net savings in total program operating costs. Figures of this magnitude lead us to expect a very high benefit-cost ratio in implementing this type of system elsewhere.

The system tested in Colorado has had a major impact on program administration because it gives adequate recognition to the high rate of change in circumstances within the recipient caseload. Implicit in the opposition to frequent reporting requirements for welfare programs is the persistent myth that AFDC caseloads are relatively static, and that receipt of AFDC is a long-term, if not a life-term, proposition for most recipients. Although serious research on welfare programs has consistently contradicted this perception, it remains the

basis for many policy decisions. In the Colorado Monthly Reporting Experiment, however, fully thirty percent of the AFDC caseload experience a change in circumstances significant enough to affect their eligibility for welfare or the amount of payment during each month. More specifically, twenty percent of the ongoing cases require a change in grant, while five percent of the cases are discontinued and another five percent are added to the caseload. This finding provides graphic evidence of the extreme fluidity in the financial needs of AFDC caseloads, and suggests that frequent reporting and rapid processing of payments are necessary if administration of the program is to be accurate and responsive.

These findings, then, are not relevant only to AFDC, but extend by implication to the administration of other programs serving low-income households such as Food Stamps and any national income maintenance program contemplated under welfare reform. The potential for improving the administration of the Food Stamp program utilizing this type of system is particularly intriguing since there is substantial similarity between the recipient populations of the AFDC and Food Stamp programs. Recent efforts to simplify the provisions of the Food Stamp program through implementation of a standard deduction would make a monthly retrospective reporting system even easier to design and implement for that program. With this type of system, moreover, the administration of the two programs could be integrated and markedly simplified through the use of a common form, comparable procedures, and compatible filing schedules. This would be a boon to those recipients who obtain benefits under both AFDC and Food Stamps since they would be faced with a single form and a single process for obtaining both types of benefits. It would also substantially reduce administrative costs for disbursing benefits to joint recipients of the two programs.

In our view, findings from the Denver Monthly Reporting Experiment are crucial to welfare reform considerations as well. Congress may soon be considering proposals for some type of national income maintenance program. It should be clear by now that a program with the noblest objectives can still fail if its administrative structure is inadequate. It is therefore very important that any administrative design take full account of the fluidity of recipient circumstances documented in the Colorado Monthly Reporting Experiment. Failure to do so would most likely lead to the same types of problems that plague the administration of AFDC, such as lack of responsiveness to recipient needs, and persistent underpayments and overpayments. Implementation of an unresponsive program would be extremely detrimental to the States and localities, as well as to recipients, for these jurisdictions would be responsible for remedying any lack of responsiveness in a national program through local programs of emergency assistance. Thus, we believe that the concepts of monthly reporting and retrospective accounting should be considered carefully in the course of drafting legislation and formulating the administrative design for a national income maintenance proposal. These concepts can help such a program to attain more closely the goals of fairness and effectiveness in which this committee has so properly expressed its concern.

○