THE NEW JERSEY GRADUATED WORK INCENTIVE EXPERIMENT

by

David N. Kershaw

and

Felicity Skidmore

July 1974

The research reported herein was performed pursuant to a subcontract between the Institute for Research on Poverty, University of Wisconsin and Mathematica, Inc., under a grant from the Office of Economic Opportunity, an agency of the United States Government, to the Institute. The research was performed by many persons both from the Institute for Research on Poverty and from Mathematica, Inc. The accuracy of the statements and interpretations made in this paper, however, are the sole responsibility of the authors.

THE NEW JERSEY GRADUATED WORK INCENTIVE EXPERIMENT[1]

by

David N. Kershaw and Felicity Skidmore

On June 30, 1967, the Office of Economic Opportunity approved a grant to the Institute for Research on Poverty of the University of Wisconsin, to undertake a controlled experiment in negative income taxation in conjunction with Mathematica, Inc., a research firm in Princeton, New Jersey. The central question the experiment hoped to address was the cost of a nation-wide guaranteed annual income as determined by the extent to which families would reduce their work effort in response to negative income tax payments. It was also expected that the experiment would provide policy makers with estimates of the administrative costs of such a program.

---

[1]This paper draws heavily from two sources: David N. Kershaw and Jerilyn Fair (Eds.), Operations, Surveys and Administration, The Final Report of the New Jersey Work Incentives Experiment, Volume IV, Institute for Research on Poverty Monograph Series, Academic Press, forthcoming; and Felicity Skidmore, "The New Jersey Graduated Work Incentive Experiment and the Decisions that Shaped It," to be published by the Brookings Institution in Joseph Pechman and Michael Timpane (Eds.), Critique of the New Jersey Experiment, forthcoming. Section 3, on the results, uses statistical material from Harold Watts and Albert Rees (Eds.), Central Labor Supply Response, Volume I, Final Report of the New Jersey Graduated Work Incentive Experiment; and Summary Report: New Jersey Graduated Work Incentive Experiment, U.S. Department of Health, Education, and Welfare, December, 1973.

The first 14 months of the grant were spent in planning, the next four years in the actual operating phase of the experiment, and the final 16 months in the analysis of the data collected and the production of the final report. The experiment cost a total of 8 million dollars, of which about one-third (2.7 million dollars) went for cash payments to the participating families.

The basic criterion for eligibility to participate in the experiment was twofold. First, the family had to contain an able-bodied male aged between 18 and 58, who was neither going to school full time, nor institutionalized, nor in the armed forces. Second, the family's normal or expected income could not be more than 150 percent of the official poverty line. Originally enrolled in the experiment were 1,216 such families--725 in the experimental groups and 491 in the control group. They were enrolled sequentially in four sites, as follows:

August, 1968 - Trenton, New Jersey
January, 1969 - Paterson-Passaic, New Jersey
June, 1969 - Jersey City, New Jersey
September, 1969 - Scranton, Pennsylvania

In October, 1969, 141 additional families in Trenton and Paterson-Passaic were added to the control group.

The operating phase of the experiment lasted three years in each site. Each family in the experimental groups filled out an Income Report Form every four weeks which formed the basis for calculating their payments. The payments were thus recalculated every four weeks, and the family received the indicated amount in two bi-weekly checks. These transfer payments were ruled by the Internal Revenue Service to be non-taxable.

A negative income tax plan is defined by a guarantee level (the level of payment the family receives if its other income is zero), and a tax or reduction rate (the rate at which the payment is reduced for each dollar of other income). Eight such negative income tax plans were finally chosen for the experiment (combinations of three tax rates and four guarantee levels), as follows:

Plan	Guarantee (Percent of Poverty Line)	Tax Rate (Percent)
I	50	30
II	50	50
III	75	30
IV	75	50
V	75	70
VI	100	50
VII	100	70
VIII	125	50
Control Group	0	0

In addition to the negative income tax transfer payments, every family receiving payments was paid a bi-weekly amount of $10 (included in their regular check) in return for sending in the Income Report Form. The controls were paid $8 a month for sending in a small card giving their current address. Every three months, an hour-long interview was administered to controls and experimentals alike, for which they were paid $5. All these other payments were considered as taxable income, unlike the negative income tax payments themselves.

Every effort was made to use the mails for sending these forms and payments back and forth, because the experiment was explicitly designed to minimize, in contrast to welfare, discretionary personal contact with the families. It was found essential, however, to have a field office at every site to deal with the filing problems that did arise.

From the beginning, the personnel administering the interviews were completely separate from those dealing with the report and payments forms. Two different names were in fact used for the two groups in an effort to underline their independence in the eyes of the families. The Payments Group was called the Council for Grants to Families (a registered trade name), and the Interviewing Branch was entitled Urban Opinion Surveys (a name now applied to an entire division of Mathematica).

The 12 regular quarterly interviews provided the main data source for the experiment. They were approximately one hour long and included two sections: a 20-minute section repeated every time on the labor force status and participation of all family members 16 years of age and older, and a 40-minute section that varied by quarter and covered at differing frequencies other kinds of economic behavior items such as expenditure and debt accumulation, plus information on health and social behavior. In addition to the quarterly interviews there were four special one-shot interviews: (1) a short screening interview, simply designed to assess eligibility for inclusion in the experiment; (2) a "pre-enrollment" interview, to collect extensive baseline data on all the families selected, before they were actually enrolled; and (3) a follow-up interview administered three months after the last transfer payment, designed to explore labor force behavior after payments had ended and to determine the families' understanding of the experiment and their reactions to the interviews and the transfer payments (for experimentals only).

1. DESIGN OF THE EXPERIMENT

The New Jersey experiment was the first large-scale controlled social experiment--that is, a field test with systematic variation of the variables of interest and a control group similar to the group receiving the stimulus in every way except for that stimulus.

Today, there are in operation several large-scale, federally funded, controlled social experiments with a basically similar approach to producing scientifically valid causal information. In 1966 and 1967, however, it was a novel idea to test economic hypotheses in this manner.

Intact Working Poor Families. Before OEO made the grant it had already been decided that the population of interest would be able-bodied working age males with dependents. Historically speaking, able-bodied males were the only people in American society who had never qualified, as a group, for non-work-conditioned public assistance. For this reason, little was known about what their work-effort behavior would be if they were to become eligible for such a transfer program. Economists had studied labor supply for many years, and over the last 30 years much empirical work had been done. But the data used had not been restricted to low-income workers and did not contain variations in income or wage rates as major as those envisaged in the negative income tax plans under consideration. Economists could not urge planners to be confident in extrapolating from minor changes affecting the general population, to what would be a major change affecting the poor. Furthermore, survey data could give no evidence on the direction of causation.

The Choice of New Jersey. The decision to concentrate on the urban Northeast was taken by OEO in early 1967. Almost everyone concerned with administrative feasibility favored a focused experiment (i.e., a sample consisting of a homogeneous group in a geographically limited area) rather than a national sample. In 1967, there was real doubt as to whether it was possible to run a social experiment at all, and a limited sample would certainly be more manageable to administer. There was no administrative expertise or experience, such as has been built up in the intervening years by Mathematica and others. Nor was it clear that plans of differing generosity could be administered simultaneously without recipients making their own invidious comparisons and refusing to co-operate. The very concept of selecting a control group--families matching the experimental group in every respect except that they received no money--without community disruption was worrying. If trouble arose, it was thought that geographical proximity of the administrators to the sites would be important in enabling the experiment to function.

The planners also wished to have a relatively homogeneous sample, another argument against an experiment with a national focus. The sample size that had emerged as statistically reasonable and financially feasible was around 1,000 families, which assumed that the sample observations could all be analyzed as one group. The existing labor-supply literature indicated differential behavior according to the type of labor-market participants and the type of labor market facing the potential worker. The decision was made to allocate first priority[1] to males with dependents in the urban industrial Northeast.

[1]It was assumed that future experiments would concentrate on other populations, other geographical locations, and other kinds of labor market.

An OEO analysis of the labor markets of five states to determine which of them approximated most nearly the overall U.S. employment rate, showed New Jersey to be a good potential site. Further arguments favoring New Jersey were twofold. First, state cooperation was promised. The concern over community disruption led the planners to put great weight on this factor. The New Jersey Department of Community Affairs, and the New Jersey Economic Policy Council promised their full support. The second argument in favor of New Jersey concerned the potential problem of overlapping welfare payments. The planners of the experiment presumed that to select an unbiased sample, and to produce usable data, it was important to avoid sites where there were competing transfer payment options. Since New Jersey had no AFDC-UP program and was not projecting one, this was an added advantage.

<u>Rules of Operation</u>. The first major task for the experiment after the grant was awarded was to draw up a set of Rules of Operation, as detailed as any potential legislation. Rules affect behavior, and a non-exhaustive set could introduce unknown and uncontrolled variation. The main definitions contained in the rules are specified below:

(a) Filing Unit. The family was chosen as the filing unit for the purposes of payment calculation and receipt of payment. The family unit was defined as the head of the family, plus his or her dependents. There had to be at least one dependent, not necessarily a child. There also had to be at least one able-bodied male between 18 and 58[1] years (not institutionalized, in school, or in the armed forces) who did not have to be the family head. The head of the family was therefore considered to be the sociological head,[2] as is also true for the Census. A dependent was defined as any blood or adopted relative living with the head, or a person not so related but living with the head and receiving no more than 30 dollars per month from income sources outside the family income.

The general intent of the family unit regulations was to replicate insofar as possible the conditions of a national program. On two specifics, however, it did not seem possible to replicate a national program--when a member leaves a family unit, and when a new member joins. There was much concern that the experimental rules be "neutral" with respect to

---

[1]The age of 58 was chosen so that none of our able-bodied males would become eligible for retirement before the end of the experiment.

[2]This decision on the family head status was not taken consistently until mid-1969, after Trenton had been in operation for a year.

incentives for family composition to change.[3] Thus, a decision was made that children leaving the family unit because they became of age could take their marginal payment with them, but <u>not</u> start a new filing unit of their own. This was to prevent the creation of a "dowry effect" (whereby children in experimental families would be more than usually marriageable for the three-year experimental period) although it was realized that, in consequence, we could probably not say much if anything about the effects of income maintenance on geographical migration. An analogous decision was made that families could admit no new members for payments purposes except natural-born children, or other children after a six-month continuous period of residence within the new family. This was to prevent experimental families from becoming artificially attractive as lodging places for relatives or friends.

When a spouse left the original household, the New Jersey regulations provided for that spouse to take the fraction of the family guarantee allotted to him or her. The guarantees for both spouses were always equal. The children's guarantees went to the spouse who took custody.

(b) Definition of Income. The basic concept of income

---

[3] The Final Report, Vol. IV, discusses in more detail the inherent ambiguity of such a concept, shows how the various income maintenance experiments differ in their regulations regarding family unit rules, and argues that how the rules treat family formation or dissolution may have critical implications for the cost and impact of a national program.

for the experiment was that used in the Internal Revenue Code. Certain additions, deductions, and exclusions were overlaid to render more comprehensive the definition of income on which payments were calculated. The purpose was to devise a measure of overall economic well-being which would enable the payments calculation to reflect adequately and uniformly differences from family to family. Additions included: annuities, pensions; prizes, awards; life insurance in excess of 1,000 dollars; gifts; alimony, court-ordered support; the rental value of public housing to the extent it exceeded rent paid, rental value of owner-occupied housing; any indirect or direct cash payments and the value of in-kind lodging from job or public[1] or private agency, including Unemployment Compensation, strike or other unemployment benefits, Social Security (Old Age, Survivors, Disability and Health Insurance) benefits, Veterans' Disability benefits, training stipends. The deductions allowed for in the experiment included: a fixed property allowance in computing income from owner-occupied housing; alimony and court-ordered support payments up to $30 a month per supported person; the cost of caring for any child

---

[1]In the original rules, welfare payments were treated as income. In 1967-68, the only major welfare program in New Jersey was AFDC. Since to be eligible for the experiment a family had to contain an able-bodied non-aged male, we expected a very minimal overlap with welfare. When a generous AFDC-UP program went into effect in January 1969, the rules had to be changed so that in any one reporting period a family could be eligible only for the experiment <u>or</u> for welfare. The rules permitted them to switch as many <u>times</u> as they liked. Treatment of welfare is discussed further below.

or incapacitated person if a member of the unit is thereby released for immediate earnings up to $80 per month for one dependent and a maximum of $120 for two or more. An annual disregard of 1,200 dollars was also allowable for aged ineligible members of the unit. Capital gains were treated as income and capital losses were counted deductible to the extent of capital gains realized during the experimental period.

(c) Accounting Period. This is the time period for which income is counted in the calculation of the transfer payment. (The accounting period for the current welfare system is generally one month. That for the positive tax system is one year.) The period for which income is included affects the degree of equity in the system, and also the degree of responsiveness to financial need. It can also be expected to affect patterns of work behavior, and therefore the transfer cost of the program. The goal of maximizing responsiveness always conflicts with the goals of producing annual equity and minimizing work disincentive and cost.

When the regulations were being drawn up, these factors were appreciated to some extent, but the conceptual and administrative problems involved in achieving annual equity without losing most of the responsiveness to need in the system were much underestimated. It was not until the experiment had been in operation in the pilot city for a year that a satisfactory system was finally achieved.

When the experiment began, two methods of maintaining annual equity were considered. The first was to make regular monthly[1] payments based on the family's income simply for the three previous months, and then reconcile payments at the end of the year. The second, instituted for only 65 families, was a 12-month moving average in which the current month's income was averaged with the preceding 11 months' income to reach an average income figure for the past year.

Much too little thought, as it turned out, had been given to the problem of recapturing overpayments, possibly because the amount of income variation that would be experienced by the families must have been implicitly underestimated, even though it was discussed at length. By the end of the first calendar year of operations in the pilot site (Trenton), which in fact only included five months of payments, substantial discrepancies between what the families had received and what they should have had coming to them on an annual basis had already appeared for 40 percent of the families. One overpayment was as high as 463 dollars. Recapturing such large sums in one installment was obviously not a possibility. It also seemed unsatisfactory to recapture out of a series of future payments, on the grounds that the responsiveness of the payments to income changes would be obscured.

---

[1]Strictly speaking, they were every 4 weeks rather than every calendar month.

The problem was finally solved in mid-1969, because the planners of the upcoming rural experiment were formulating their regulations and were having to tackle explicitly the problem of self-employed farmers whose incomes and outflows are very bunched throughout the year. They devised a carry-over method of accounting under which payments made to a family were based not only on the average income of the last three months, but also on any income in excess of the "break-even" amount earned in any of the preceding twelve months. As long as a carryover sum existed, the payment was zero. When all the carryover was used up and the moving average was under the breakeven point, payments were resumed. The carryover was used up in a "first in first out" fashion. That is, carryover generated in Month 1 was used up before that generated in Month 2, and so forth. If the carryover was not used up within 12 months, it was dropped off.

Table 1 shows how the irregular earner would fare on a monthly accounting system with the carryover provision. As the table shows, total income (earnings plus payments) has been substantially smoothed and total payments for the year add to the correct annual amount.

Overlapping Tax Rates of Other Programs. An important issue was that of maintaining control over the tax parameter. It was considered impossible to create a situation where the families had no contact with or participation in other tax and transfer mechanisms. The only feasible alternative,

TABLE 1

TREATMENT OF THE IRREGULAR EARNER UNDER A MONTHLY ACCOUNTING SYSTEM WITH CARRYOVER

Month	Earnings	Accumulated Carryover Sum	Net Income (Earnings + Carryover Sum)	Newly Added Into Carryover	Payment	Total Y
January	600	0	600	100	0	600
February	800	100	900	300	0	800
March	200	400	600	0	0	200
April	0	100	100	0	200	200
May	0	0	0	0	250	250
June	300	0	300	0	100	400
July	500	0	500	0	0	500
August	600	0	600	100	0	600
September	600	100	700	100	0	600
October	0	200	200	0	150	150
November	200	0	200	0	150	350
December	200	0	200	0	150	350
Total	4,000				1,000	

Note: Annual breakeven is $6,000 and monthly breakeven is, therefore, $500.

therefore, was to decide which other tax rates were important to control for and then formulate a means of counteracting their effects. By the following reasoning, the issue narrowed fairly readily to the question of the positive income tax.

First, a decision was taken not to concern ourselves with those programs which could be expected to remain in operation even if a negative income tax were adopted nationally, such as the sales tax. That is, the objective of interest was the labor-force response of people on experimental plans in comparison (within the real-world context that could be expected to surround a national program) with that of the control group.

Work-conditioned tax rates, on the other hand, were recognized to present a major problem. Any tax rates that might affect the work effort of controls and experimentals differentially would have to be controlled for in order to observe the experimentally induced effects.

The welfare system was an obvious candidate on both counts. It could be expected to vanish with the implementation of a national program, and it was a work-conditioned program. As noted above, the planners had picked New Jersey in part because, with the population of interest being families with an able-bodied non-aged male, the absence of

AFDC-UP meant that the problem of a welfare overlap disappeared.[1] It was realized that during the course of the experiment certain families would in all likelihood become female-headed and thus potentially eligible for AFDC; but this was assumed to be a small number, and could therefore be handled adequately simply by including welfare payments in the experimental definition of taxable income. Originally, Medicaid and Food Stamps were considered equally minor--Medicaid because it was tied to AFDC eligibility, and Food Stamps because it was a very small program in New Jersey at the beginning, and indeed for a large part of the experiment.[2]

Social Security was rationalized as no problem because it was not at that time envisaged as disappearing with welfare reform. Although a work-conditioned program, it was considered to be part of the status quo background for controls and experimentals, and unlikely, therefore, to affect work choices.

That left the positive income tax system, which was handled (not entirely satisfactorily) in the following way.

---

[1]See below for a more detailed discussion of the introduction of AFDC-UP and experimental efforts to deal with it.

[2]An attempt was made to negotiate Medicaid eligibility for experimental families, but it was unsuccessful, unlike the subsequent income maintenance experiments in which families are eligible for welfare medical benefits even though they don't receive welfare cash benefits.

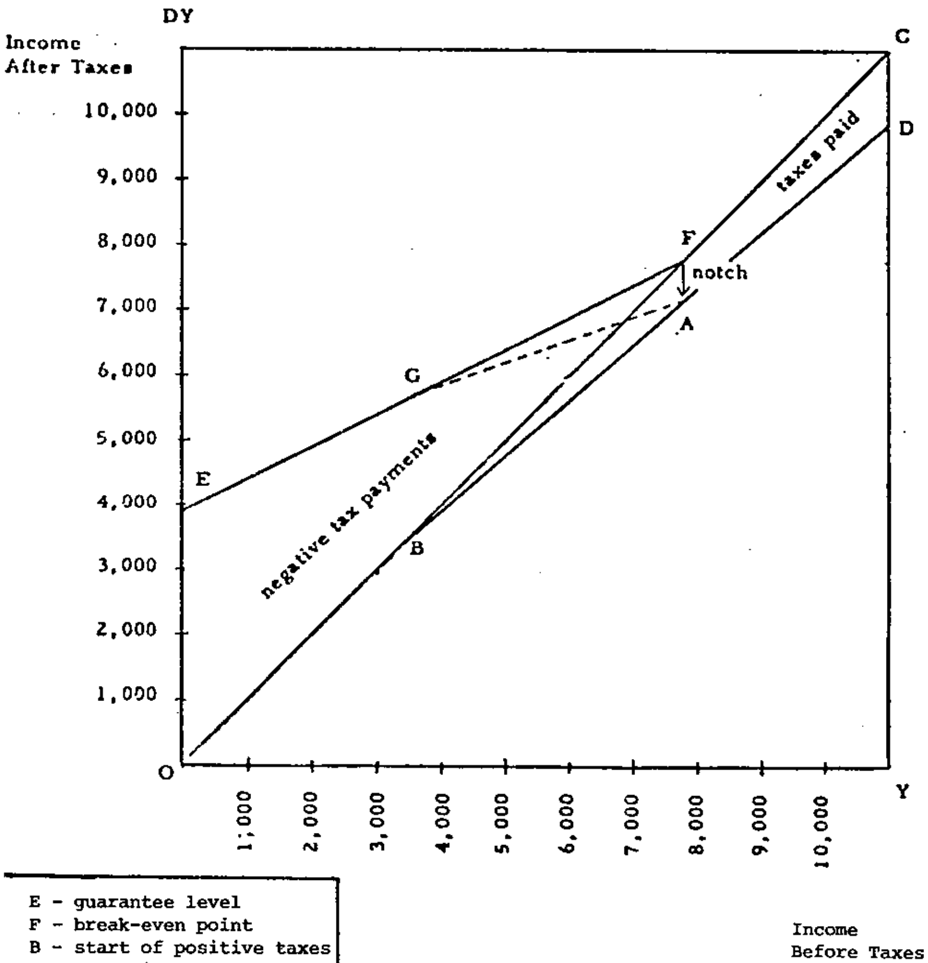
Because the maximum income level at which no positive tax was paid was lower than the income level (breakeven point) at which experimental payments were reduced to zero, superimposing one system on the other with no adjustment produced a range in which experimental families faced the experimental benefit reduction rate plus the positive tax rate. In addition, if their income increased above the breakeven point for their plan, they faced a tax rate "notch" which produced a <u>fall</u> in their disposable income. A decision was made that families would be reimbursed in full or in part for positive taxes paid so as to maintain a constant tax rate (dictated by the experimental plan they were assigned to) up to the "tax breakeven point," defined as the point where income minus federal tax paid was exactly equal to the breakeven point of their experimental plan. Families paying no positive tax were thus unaffected; families with incomes below the breakeven point of their experimental plan had all taxes reimbursed; families with incomes over the breakeven point for their experimental plan but not as high as their "tax breakeven point" were partially reimbursed. See Figure 1 for a graphical representation.

Conceptually, this was a straightforward solution. From an experimental point of view, however, it turned out most unsatisfactorily because the administrative procedures adopted for the rebate produced a major time lag between the time the families paid their positive taxes and the date of their rebate check. Rebates were calculated on the basis

Figure 1

COMBINED POSITIVE AND NEGATIVE TAX SYSTEMS, UNADJUSTED

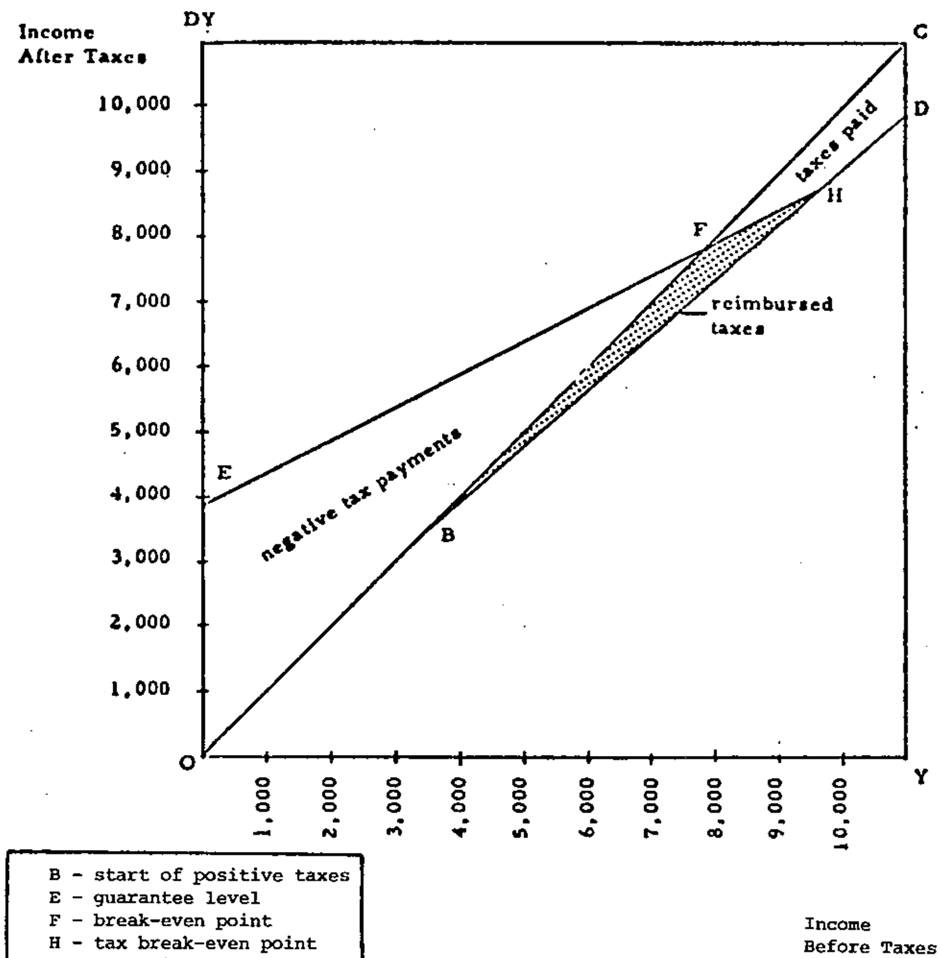
(Family of Four)



E - guarantee level
F - break-even point
B - start of positive taxes
EGFC - gross income
EFAD - disposable income, 1
EGAD - disposable income, 2

COMBINED POSITIVE AND NEGATIVE TAX SYSTEMS
ADJUSTED BY TAX REBATE SYSTEM

(Family of Four)



B - start of positive taxes
E - guarantee level
F - break-even point
H - tax break-even point
EFC - gross income
EFHD - disposable income

of W-2 forms and 1040 tax forms which were requested from the families. The families were for the most part quite willing to hand these over. But using the 1040 forms meant that taxes paid in January, for instance, could not be rebated until more than a year from the following April when the 1040 forms were due.[1]

<u>Composition of the Sample</u>. The decision to restrict the sample of interest to families containing an able-bodied male and at least one dependent has been discussed above, as has the decision to restrict the geographical area to urban industrial sites in New Jersey. Here we shall discuss the decision to restrict the sample to families with normal family income of not more than 150 percent of the poverty line.

From a research point of view, the higher the income[2]

---

[1]In the Seattle-Denver experiment tax rebates are made concurrently with the payments on an estimation basis. A final reconciliation then has to be made with a lag as long as New Jersey's but the amounts are much more significant.

[2]The concept of income being discussed was not current income, but rather some measure of normal, expected, or permanent income, to correct for the possible truncation bias that might be introduced by having people with incomes that were unusually low (for them), over-represented, and people with incomes higher than their usual excluded. For the final analysis two statistically rather sophisticated measures of normal income were formulated. It was originally hoped that the same could be done for the normal income measure on which the sample was selected and stratified. The exigencies of the field work schedule and data processing facilities were such, however, that the estimates for eligibility purposes were made by inspection of the screening and pre-enrollment questionnaire to see whether certain features of the family (education and training of the wife, for instance, if she wasn't working) would lead to the presumption that their normal income was probably significantly higher or lower than the average given for last year.

cutoff (within reason) the better the estimates of cost and work effort would be. Two general considerations that militated against raising it further were (1) the concern of those close to the field operations that the payment amounts not be insignificant for a large group of sample families, and (2) the concern in OEO that they not be in the position of funding a program addressed to the non-poor.

The decision as to the income cutoff had a major unfortunate consequence that was not perceived as the decision was being made. It had the necessary effect of leaving working wives very under-represented in the sample. The percentage of wives who remained in intact families and worked regularly throughout the experiment was only 15 percent (although 40 percent worked at some time) leaving an uncomfortably small sample for analysis.[1]

Specification of Experimental Variables. The experimental variables were the guarantee levels and the tax rates. The guarantee levels (adjusted for family size) were specified as percentages of the poverty line. In fact, the official poverty lines were altered for the experiment, because it was felt that there was no analytical justification for using nutritional requirements alone as the basis of need (as the Social Security poverty levels did), especially regarding the variation in need by family size. The poverty levels used for the first year of experiment are given below,

---

[1]It was particularly uncomfortable because of the fact, also unanticipated, that the three ethnic groups had to be analyzed separately.

along with the 1967 Social Security Administration (SSA) index for comparison. They were increased in July of every year by the percentage change in the Consumer Price Index.

Poverty Levels by Family Size

	Experiment (First Year of Operation)	SSA Poverty Line (1967)
2 persons	2,000	2,130
3 persons	2,750	2,610
4 persons	3,300	3,335
5 persons	3,700	3,930
6 persons	4,050	4,410
7 persons	4,350	4,925
8+ persons	4,600	5,440

The combinations of guarantee levels and tax rates originally chosen formed the following policy space:

Guarantee (% of pov. line)	Tax Rate (percent) 30	50	70
50	X	X	
75	X	X	X
100		X	X

The field operations began in the summer of 1968. The yield of eligible families was roughly half what had been expected, and only about 30 percent of the eligibles had incomes actually below the poverty line. This meant, in turn, that payments to the families would run about 40 percent below the target set earlier in 1968 (an average per family of $1,500 a year). In the Fall of 1968 another development aggravated the problem. The State of New Jersey decided to introduce, as of January 1, 1969, an AFDC-UP program with the following support standard:

New Jersey AFDC Standard

2 persons	$2,300
3 persons	3,175
4 persons	3,800
5 persons	4,250
6 persons	4,650
7 persons	5,000
8+ persons	5,300

These levels were generous enough to dominate the two lowest plans over almost the entire eligible income range, and most of the others over parts of the range. The certainty that a UP program would be introduced with one of the highest support standards in the country persuaded OEO that changes had to be made. A new guarantee of 125 was added to the policy space, to be combined with the 50 percent tax rate.

Allocation of the Sample. Divergence between Mathematica and the Institute for Research on Poverty as to how the sample should be allocated among the experimental began to appear as early as the Spring of 1968; disagreement continued through the actual assignment of families in both Trenton and Paterson-Passaic, with subsequent consequences for site, plan, and ethnicity confounding in the data; and was finally settled in June of 1969, just as Jersey City was being enrolled.

The objective of the experiment was agreed to be estimating the cost of a national program (or at least the cost of covering urban wage earners in the Northeast). It was also agreed that central to this objective was how to distribute eligible families among the various experimental cells so as to maximize the information obtained subject to the budget constraint imposed by OEO.

As 1968 progressed, alternative allocation models and calculations were made showing severalfold differences in cost per unit of information according to the way the sample was allocated. The decision was therefore made to allocate families in Trenton and Paterson-Passaic in such a way that no cell received more than the minimum number of families shown by those calculations as a possible outcome.[1]

It became rapidly apparent that the allocation model being formulated at the Institute was leaving many more people without positive payments (that is, either on plans where

---

[1]There were 21 possible cells at this stage, because the sample was stratified according to three normal-income strata: 0-99 percent of poverty, 100-124 percent, and 125-150 percent.

they were above their breakeven or in the control group) than anyone had envisaged, or than was comfortable from the point of view of the field staff. The experiment was being somewhat less than enthusiastically received by certain community groups, and the field staff were only able to hold off militants in Paterson and Passaic by explaining that the experiment would bring money into the community. High and continuing attrition was also being encountered because of a lack of financial incentive to continue. An additional concern was expressed, especially by Mathematica sociologists. With a large proportion of the sample either above their breakeven points or receiving very small payments, certain research areas where the distinction of interest was whether or not payments were received rather than the varying size of the payments stood to suffer from a paucity of data on the payments group.

The Wisconsin position was that the central behavioral response was with work effort. The model should therefore be designed to optimize information on that, with other responses estimated as well as possible from the data that emerged; any dissatisfaction with the resulting allocation should lead to changes in the assumptions made rather than a jettisoning of the model itself. The model contained the following constituent parts.

(1) The budget constraint dictated that the cost of the observations be weighed in any assessment of the information they would be expected to contribute. (For the cost of every

low-income stratum family on the 125/50 plan, 33 high-income families could be assigned to the 50/50 or 75/70 plan.) (2) Policy considerations implied that some plans were more important in terms of national options than others, so that each plan should be assigned a policy weight. (3) There was an income level above which a negative income tax plan could be expected to have no effect on the work behavior of the family. This level would be above the breakeven point, and could be expected to increase as the tax rate increased. (4) Attrition would be a function of plan generosity, and the lower the expected payment the higher would be the expected attrition. (5) Variance in earnings behavior would be higher for controls than for experimentals because, for the same labor market conditions, the effective wage change would be smaller for those receiving benefits. (6) The response function would be continuous across cells. Each cell, in other words, was not assumed to be the only source from which data could be obtained about behavior in that cell.

None of these specifications were objected to as such. The disagreement can perhaps be captured by saying that the Wisconsin model took as the objective of the experiment the measurement of the cost of a national program in the strictest sense. Other objectives were not considered as reasons for modifying the model. Mathematica, on the other hand, particularly by the Spring of 1969 after nine months of partial operation of the experiment, saw other objectives as independently important. (1) Studying the feasibility of running a self-administrable transfer program was important; excessive attrition and few large-payment families would make this less possible. (2) The sociological analysis would suffer if only a small number got payments large enough to make a substantial financial impact on the families. (3) The model would have to be simple enough to be able to explain to the public at the end of the experiment; not to win public acceptance of the outcome would be to fail. (4) The feasibility of field operations depended on community acceptance of the experiment, which in turn depended on high enough payment levels.

The design controversy was finally referred to outside arbitration. Both sides agreed to James Tobin as the arbiter and to accept his judgment as final, and took him their arguments. Mathematica formulated a compromise plan suggesting the cells be divided into five regions in terms of payment generosity (see Fig. 2). They argued that at the end of the experiment there should be a minimum number of families in each income stratum in each payment region. Tobin decided to solve the problem by working his own best assumptions through the Wisconsin model, and making certain judgmental deviations.

In particular, he tackled the attrition problem directly by raising the flat payments to the experimental families for filling in their Income Report Forms from $2.50 every two weeks to $10.00 every two weeks; and to institute a payment of $8.00 a month for the control families (previously paid only $5.00 for every quarterly interview).

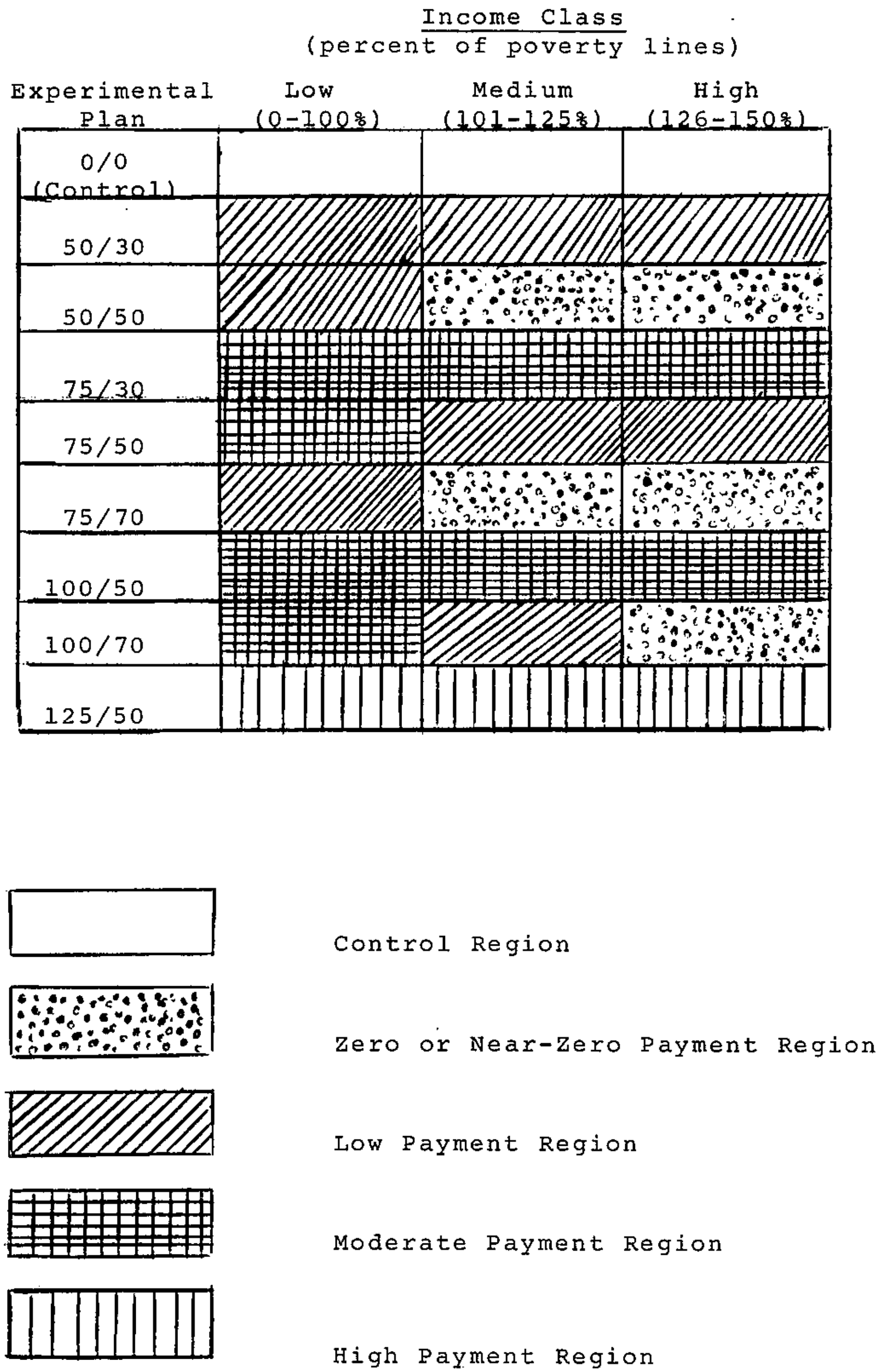


Figure 2: PAYMENT REGIONS--MATHEMATICA PROPOSAL

The resulting distribution allocated nine more poor (i.e., lowest stratum) families to the 125/50 plan than the model indicated. This violated the OEO constraint of 33 percent of the transfer budget by using up 36 percent. OEO acquiesced. The Tobin allocation is given in Table 2.

The only administratively difficult consequence of the Tobin judgment was that about 150 new controls were to be added to the sites already in operation (Trenton, Paterson, and Passaic). Those sites had to be rescreened, and the eligibles found who had not already been assigned the first time around were given a special make-up pre-enrollment interview.

2. THE FIELD OPERATIONS

The actual fielding of the New Jersey experiment turned out to be more time and budget consuming than anyone dreamed of in 1967. The main difficulty envisaged before the fact was possible community disruption from assigning people to plans of different generosity and some to no plan at all. This did not turn out to be a problem, although certain community militants tried for a time to get the experiment out of Paterson and Passaic altogether. The cooperation of New Jersey State and local authorities, on the other hand, was seen to be an important ingredient in eventual success, whereas in fact a major field problem came about because of the

TABLE 2

TOBIN ALLOCATION MODEL

(ADDITIONAL "LAGGED" ACCOUNTING GROUP FAMILIES

GIVEN IN PARENTHESES)

Plan	Treatment (% guarantee/ tax rate)	Income Stratum Low	Medium	High	Total
A	50/30	5	31	12	48
B	50/50	29	37	5	71
C	75/30	30(5)	14	50(5)	94(10)
D	75/50	5	57(10)	36(10)	98(20)
E	75/70	13(5)	51(15)	0	64(20)
F	100/50	22	34(5)	20	76(5)
G	100/70	11	26	33(10)	70(10)
H	125/50	50	8	80	138
Total Experimental		165(10)	258(30)	236(25)	659(65)
I	Control	238	165	247	650
Total Sample		403(10)	423(30)	483(25)	1309(65)[1]

[1]This differs from the 1357 total because of a sampling shortfall. It was only possible, for example, to find 141 new controls in the cities that were already in operation.

very adverse attitudes and actions of the welfare authorities. As far as state cooperation in general was concerned, it turned out to be virtually nonexistent, and not (except in the case of the welfare overlap) very necessary.

Site Selection. The reasons for selecting New Jersey have been described above. Early on in the experiment it was decided to enroll sites sequentially, due in part to the difficulty of getting field operations underway, but also due to the reasonable view that the first site could be treated as a pilot site from which lessons could be learned for the other site or sites. The main reason given for wanting more than one city was to test the effects of differences in industrial composition and labor-market tightness. As it turned out, we could not find enough eligibles in four New Jersey cities, and in fact even went outside New Jersey to complete the sample.

Trenton was chosen as the first site because of its proximity to Princeton and Mathematica, because of the cooperative attitude of United Progress Incorporated of Trenton, and because it was the seat of the State government. The second and third were Paterson-Passaic and Jersey City.

As soon as the Trenton sample was complete it was realized that the ethnic balance would be a problem. The 1960 Census showed blacks as only 22.4 percent of the Trenton population. When the full tally of the eligibles was in, however, 66 percent of the sample was black, 16 percent was white, and 18 percent was Puerto Rican. The tally when the

Paterson-Passaic sample was complete was not much better, with 40 percent black, 50 percent Puerto Rican, and only 10 percent white. In Jersey City again we found a high proportion of Puerto Ricans. Even stratifying them out (that is, taking every second Puerto Rican family), the final sample distribution in Jersey City was 51 percent black, 13 percent white, and 36 percent Puerto Rican.[1]

In the face of these results there had to be an explicit decision made to go to a site where the population could be counted on to be predominantly white. There were no other low-income areas in New Jersey where there could even be the presumption that efficient sampling could produce the requisite number of whites. The decision was made (after preliminary sampling) to go to Scranton, Pennsylvania. The overwhelming majority of those sampled did indeed turn out to be white. The experiment now had an ethnically balanced sample (roughly one-third for each of the three major ethnic groups) but since the whites were predominantly in Scranton, the ethnic groups were not balanced within the sites.

Selecting the Sample. After the sites were selected, target sampling areas within the cities were ascertained. The decision was originally made, encouraged by OEO, that sampling not be restricted solely to tracts identified by the 1960 Census as poverty areas, on the grounds that an important group

---

[1] The 1960 Census had shown Paterson as 15 percent black and Jersey City as low as 13 percent black.

of low-income families--i.e., those who chose not to live in patently poor districts--might otherwise be missed. After the Trenton experience it was clear that sampling outside the 1960 so-called poverty tracts was yielding practically nothing, and that even within the poverty tracts sample areas had to be redrawn by staff members who drove down every street. By the time interviewers were sent into the field, supervisors knew the various ethnic areas, which blocks would not have dwellings, and which areas were dangerous. This made sampling more efficient and more accurate, since both structural and socio-economic changes in urban areas since the 1960 Census turned out to be frequent and usually impossible to identify from existing data sources.

Table 3 summarizes the sampling procedures which were used to select the sample. Since this was a multi-step and complex procedure, each of the steps given in Table 3 are described below.

(a) Listing. Listing consisted of noting the address of each housing unit in randomly selected blocks within the target areas of the sample cities. These target areas were widely drawn contiguous segments of the city parts of the SMSAs, located and defined by using 1960 Census income data modified by on-site inspection. Once the target area was drawn, blocks within the areas were randomly selected at a sampling rate based on projected eligibility needs. As a practical matter, virtually all blocks within the target areas were selected for listing, since as many eligible families as

TABLE 3

RESULTS OF THE SAMPLING PROCESS

Interview Stage	Trenton	Paterson/ Passaic	Jersey City	Scranton	Total
A. Listing					
Housing Units Listed	3,530	14,781	18,002	12,334	48,647
B. Screening					
Interviews Attempted	3,497	14,781	18,002	12,334	48,614
Housing Unit Vacant	270	729	2,668	1,225	4,892
Family Never Home	400	2,954	3,997	1,063	8,414
Refused Interview	400	2,449	3,531	1,578	7,953
Completed Interview	2,427	8,649	7,806	8,468	27,350
C. Screening Eligibility Determination					
Ineligible for Pre-Enrollment	2,200	7,661	6,567	7,798	24,226
Eligible For Pre-Enrollment	227	988	1,239	670	3,124
D. Pre-Enrollment					
Interviews Attempted	227	913	1,143	670	2,953
Moved, Could Not Locate	18	177	149	57	401
Refused Interview	22	44	88	57	211
Completed Interview	187	692	906	556	2,341
E. Enrollment Eligibility Determination					
Ineligible for Enrollment	50	246	358	174	828
Eligible for Enrollment	137	446	548	382	1,513
F. Enrollment					
Attempts	137	446	413	320	1,316
Moved, Could Not Find	2	21	7	0	30
No Longer Eligible	0	3	5	0	8
Refused Enrollment	9	40	11	2	62
Enrolled	126	382	390	318	1,216

possible were wanted, given difficulty in filling the sample. Once a block was selected, all dwelling units on that block were listed on a listing sheet and then transferred to "dwelling unit cards." Thus, for each of the 48,865 housing units selected within the target areas of the four SMSAs, a single dwelling unit card existed. This card was then used both as an administrative aid in the interviewing process and as a record of the final status of each housing unit.

(b) Screening. The screening interview was designed as a first cut at eligibility. Its primary purpose was to increase efficiency by cutting down the number of families who would receive the longer and more costly pre-enrollment interview. As the table indicates, a small number of dwelling unit cards were held in reserve and never used, and a total of 48,614 screening interviews were attempted. A total of 4,892 housing units were found to be vacant and thus no contact could be expected. This left a potential of 43,722 households. Of these, interviews were completed with 27,358 or 62.6 percent, leaving 8,414 (19.2 percent) who were never found home (after five tries) and 7,958 (18.2 percent) who refused to speak with the interviewer. A standard four "call-back" technique was used, varying the time-of-day and day-of-the-week of each of the tries. In Jersey City, where a low initial yield caused concern about bias, eight call-backs were used.

(c) Screening Eligibility Determination. Among the completed screening instruments, 3,124 (11.4 percent) contained data on families considered eligible. The screening interview was brief and was not expected to determine eligibility with the precision of the longer instrument, but it eliminated those who were obviously not eligible.

(d) Pre-enrollment Interview. A small number of families were held in reserve, even though eligible on the basis of the screening instrument, for use if needed subsequently. Most of these were Spanish-speaking families (in Paterson/Passaic and Jersey City as shown) who were over-represented in the sample. Of the pre-enrollment interviews attempted, 2,341 (79.3 percent) were completed. The higher pre-enrollment completion rate can be explained as a result of having weeded out the "chronic refusers" on the screening interview. Nevertheless, 6.9 percent of those contacted did not agree to submit to the longer pre-enrollment interview and 13.6 percent could not be located even though they had taken a screening interview.

(e) Enrollment eligibility determination. Of the 2,341 pre-enrollment interviews completed, 1,513 (64.6 percent) contained data on families who were still considered eligible. Most of the reduction in eligible families came from the more detailed, and precise, income information contained on the longer instrument. As a general rule, the more detailed and probing the income questions are in an interview, the more income is found.[1]

---

[1]For a more detailed discussion of this, see D.L. Bawden and D.N. Kershaw, "Problems in Income Reporting and Accounting,"

(f) Enrollment. As before, a number of eligible families were not released for enrollment attempts. In Jersey City the ethnicity oversampling problem was again the reason, and in Scranton it was clear that the sample could be filled without using all eligibles from pre-enrollment. The reserve group in Scranton (62 families) was randomly eliminated from the potential enrollment group; 1,316 enrollments were attempted, resulting in a final sample of 1,216 (the enrollment completion rate again showing a tendency to climb over the previous pre-enrollment completion rate to 92.4 percent). As indicated elsewhere, 141 control group families were subsequently enrolled in Trenton and Paterson/Passaic, bringing the total final sample to 1,357.

As Table 3 indicates, then, the target households (those containing residents) numbered 43,722 and the number of families finally enrolled was 1,216, or 2.8 percent of the households sampled. The question thus arises as to what kind of bias may have been introduced by the elimination of so many families, not only for reasons of eligibility, but because of refusals and failure to be at home when an interviewer called.

Orr, Hollister and Lefcowitz, Income Maintenance, Markham, 1971, and D. N. Kershaw, "Administrative Issues in Establishing and Operating a National Cash Assistance Program," Joint Economic Committee, Congress of the United States, Studies in Public Welfare, Paper No. 5 (Part 3). U.S. Government Printing Office, March 12, 1973.

Table 4 measures the extent to which the final experimental sample was a faithful representation of the families fitting the characteristics sought in the target cities, as indicated by a comparison with 1970 Census statistics. The first entry shows Census data on the percentage of non-aged, male-headed families, below 1.5 times the 1968 poverty line, regardless of whether or not they are "in the labor force." Those omitted include disabled (eliminated by the experiment) and others who simply do not want to work (not eliminated by the experiment). Because this number eliminated no heads on disability grounds, it clearly overstates the eligibility rate for the experiment. The second entry shows Census data on non-aged, male-headed families with the head in the labor force who are below 1.5 times the 1968 poverty line. This understates eligibility for the experiment by eliminating all those families who, while otherwise eligible, did not have a male head in the labor force. The true Census eligibility rates applying the experimental parameters, therefore, fall within the range represented by Rows 1 and 2.

Row 3 represents the eligibility rate based on the experiment screening. It is derived simply by dividing the families "eligible for pre-enrollment" by the total completions in screening. Row 4 shows eligibility based on the pre-enrollment interview. This is a more precise measure, using the detailed pre-enrollment income questions. It is derived by factoring up the families eligible for enrollment (1,513) by the ratio of the pre-enrollments attempted to the pre-enrollments

TABLE 4

CENSUS-EXPERIMENT COMPARISON

	Trenton	Paterson/ Passaic	Jersey City	Scranton	Total
A. Census					
All non-aged, male headed families, below 1.5 percent	7.0 (1)	9.7	9.0	8.2	8.6
All non-aged, male headed families, in the labor force, below 1.5 percent	5.2	7.2	6.4	4.8	6.1
B. Experiment					
Eligibility rate based on screening	9.4(8.2,10.6) (2)	11.4(10.6,12.1)	15.9(15.1,16.7)	7.9(7.3,8.5)	11.4(11.0,11.8
Eligibility rate based on pre-enrollment	6.8(5.8,7.8)	7.4(6.8,8.0)	9.6(8.9,10.3)	5.5(5.0,6.0)	7.4(7.1,7.7)

(1) Census estimates were derived by factoring-up numbers on poverty families to include all these under 1.5 times poverty, eliminating single individuals and female-headed families, and adjusting the Census age range for heads (14-64) to conform to that of the experiment (18-58).

(2) Numbers in parentheses are upper and lower 95% confidence limits for the above percentages.

completed (to correct for moves, refusals and those eliminated). *A priori* one would expect the screening eligibility (Row 3) to fall between the bounds provided by the Census, Rows 1 and 2, and the more precise pre-enrollment eligibility (Row 4) to fall near the bottom or below that range.

As Table 4 shows, this is exactly what happened in Scranton and Trenton, and close to what occurred in Jersey City and Paterson/Passaic. As the Scranton column indicates, the screening eligibility from the experiment survey falls in the top of the range (numbers in parentheses are upper and lower 95 percent confidence limits for the observed percentages) and the pre-enrollment eligibility falls at the very bottom of the range. Trenton, while slightly above in the screening eligibility, conforms quite well. Two things may explain the differences in the comparisons for Jersey City and Paterson/Passaic. In the first place, the screening instrument may have been too imprecise on income for good comparisons (that is, being very brief it over-included families relative to the Census instrument). A second explanation is that the experiment, being very attentive to problems of bias introduced by traditional survey methods which miss black families in cities (and for which the Census has been criticized) simply found more families on the screening in predominantly black areas (Paterson/Passaic and Jersey City). The fact that this is not the case in heavily black Trenton may be accounted for by the fact that Trenton was a Census pre-test site and extra care may have been taken to avoid the black undercount problem.

Interviews. The regular survey instruments used in the experiment consisted of a "core" of labor-force questions which were repeated at each quarter, and a rotating group of questions, some of which appeared on an annual basis, some of which appeared semi-annually, a few of which appeared only at the beginning and end of the experiment, and some which were "once-only" topics. These questionnaires covered a wide range of economic and sociological topics.

It was originally assumed that a professional survey organization would carry out the interviewing. Bids were solicited, all of which were influenced too much by the customary presentation of competitive bids and not enough by the unique requirements of the experiment. The best bid was accepted but it soon became clear that unusual procedures were necessary which were not forthcoming. Mathematica, therefore, made the decision to conduct its own interviewing--employing tighter controls, longer and more in-depth training, and a much greater stress on minority staff and interviewing personnel than the usual survey firm.

Administration of each interview wave took from four to six weeks in each site, and from 15-20 interviewers. High interviewer turnover necessitated continuous hiring and training, because interviewing in urban areas turned out to demand a substantial commitment on the part of the interviewer. There were instances of assaults and robberies, which were no doubt increased because the exigencies of collecting earnings data required interviews to be undertaken when the wage-earner was most likely to be at home, after dark.

In the development of all interviews it was learned on-the-job that questions had to be constructed with special attention paid to the particular socio-economic characteristics of the sample. Many of the questions used had been developed elsewhere, but pretesting and field experience made it clear that we had to reduce many polysyllabic words to simpler ones. While conserving original wording is strongly supported by survey researchers, the percent of families with limited vocabularies necessitated certain changes. If a word proved particularly difficult to a family the interviewer was forced to paraphrase. Preserving internal consistency required generally usable wording at the outset to prevent interviewers from having to paraphase individually. A particular problem was created by the Puerto Ricans, 25 percent of whom had to be given interviews in Spanish even though sociological and attitudinal questions seemed to translate badly.

A major unexpected lesson learned from the interviewing branch of the study relates to the central labor-force and income data. The original decision was to use questions from the Current Population Survey. A set of well-known standard questions, it was assumed, would minimize the risk of unusable data and maximize the likelihood of being able to make rational comparisons. By mid-1969, however, these questions were seen to be seriously inadequate because they did not provide anything but a one-week snapshot out of every quarter. There was, thus, no complete and continuous work and income history for comparison with the Income Report Forms. They also occasionally picked up atypical weeks (holidays) for large parts of the sample.

In July 1969, therefore, the labor-force "core" of the questionnaires was amplified by expanding the "last week" set to "last month by weeks." The definition of "last month" was associated with pay periods, which simplified the interview, increased accuracy, and aided the matching of interview income data with Income Report Form data.

Welfare. As described above, on January 1, 1969, New Jersey introduced a generous AFDC-UP program for which most of the experimental sample were eligible and which dominated most of the experimental plans over at least some of the relevant income range. Throughout the Fall of 1968, alternatives were discussed as to how to treat this problem. As we have seen, a new experimental guarantee of 125 percent of poverty was added to the policy space to forestall defections to welfare. In addition, it was decided that, for all the sites except Trenton, the Rules of Operation would have to be changed so that experimental families would have to decide every payment period whether to accept welfare or experimental payments. They were to remain in the sample for the purposes of sending in Income Report Forms and answering the quarterly interviews, and they could change back and forth as many times as they liked between welfare and experimental payments.

The Trenton sample was to be allowed to continue as it had started, reporting any welfare payments to the experiment as income. In November 1969, the Mercer County prosecutor's office began investigating overlapping payments, and subpoenaed 14 family records. Mathematica moved to quash the subpoenas, and after weeks of negotiating with the prosecutor and actually monetary haggling with Mercer County, the experiment agreed to pay back the approximately 20,000 dollars of overpayments that it was established the families had received.[1] The Trenton rules had to be changed to conform to the rest of the cities, and a system of quarterly checks with the welfare departments for overlap was instituted.

Nixon's Family Assistance Plan. In 1967 it was inconceivable to everyone that political reality would overtake the experiment. In August 1969, however, the Family Assistance Plan, with a negative income tax component built into it, was announced as a welfare measure by the Administration. The experiment was the only potential source of empirical data relevant to the proposed legislation. In January, 1970, representatives from the experiment were called to testify before the Ways and Means Committee. They replied to questions in general terms, but did not have any hard data for the Committee and had not in fact decided whether giving preliminary data would be indicated even if available. After their testimony, the Poverty Institute, Mathematica, and OEO

[1]Later, Passaic County attempted also to collect some money from the experiment, but since the rules for the Passaic sample were unambiguous, Mathematica was able to threaten a countersuit and the case was dropped.

all felt that preliminary results, even if it meant hand tabulating, should be released. On February 18, therefore, Preliminary Findings of the New Jersey Work Incentives Experiment was issued by OEO. This decision let the experiment in for extensive and critical examination by the General Accounting Office (GAO). It set in motion an attempt on the part of the Senate Finance Committee to obtain confidential files on individual families. Lawyers at OEO gave the opinion that even though such files were confidential, the experiment would not be able to resist a Congressional subpoena. It also let the research staff of the experiment in for severe professional criticism for allowing themselves to be used as a fig leaf for administration advocacy.

In testimony before the Senate Finance Committee on August 18, 1970, the experiment staff, OEO, and the GAO were all called to testify. The GAO was, although critical of the preliminary report itself, supportive of the experiment, and the threat of a Congressional subpoena dissolved. The Institute, meantime, had issued its own appropriately tentative version of the preliminary results for the research community.

The general lessons to be learned are these: Congressional interest and concern pose a sensitive issue. It is helpful to establish a relationship with the GAO when the experiment is launched. When an experiment becomes as relevant to the policy-making process as New Jersey did, it requires a difficult but essential balance between the interests of research and policy. When results are obtained they should be published with suitable qualifications, and the GAO should see them before they are released.

The approach subsequently taken with H.R.1 ("Social Security Amendments of 1971") is probably the best course. The research staff disclaimed for the experiment any central relevance to the legislation but indicated to both the House Ways and Means Committee and the Senate Finance Committee that they were willing to act as a resource for questions which fell within the scope of the experimental experience. The Ways and Means Committee took advantage of this offer. Neither OEO nor the experiment staff supported or criticized the legislation, only supplying answers to specific questions raised.

Termination. The final major decision to be made regarding the operation of the experiment concerned how to terminate payments to the families. At the start of the experiment the families were explicitly told the duration of the payment phase, and they were given a wallet-sized card with the termination date printed on it. After their enrollment, however, no new mention was made of the final payment date. Ethical concerns had been expressed from the beginning about creating possible hardship by accustoming families to income augmentation and then withdrawing it. As the time for the last payment in Trenton approached, discussion intensified as to our ethical responsibilities with regard to the approaching end to the payments.

It was finally decided that--since what we did in

Trenton should be regarded as a pilot, not only for the rest of our families, but also for Iowa, North Carolina, Seattle, Denver and Gary--we had to proceed in an experimentally justified way (i.e., with no extra warning before the final interview and no gradual tapering off of payments) so that we could get some real information on family and public reactions. The last quarterly was given before the last payment; and the families were reminded of the last payment date once, just after the final quarterly interview.

The field offices remained open as referral agencies for any hardship cases, but not a single family came to them for assistance. The only other data that give information on the experience and attitudes of the families after the end of the experiment are from a questionnaire administered three months after the last payment, designed to measure responses to termination. The families reacted rationally in the sense that the relative importance of the payments to their total income affected how much thought they have given beforehand to the termination of those payments and what preparations they made for it. There were, however, no major differences in terms of the special budgeting, emergency, or medical needs they said they had actually encountered during the payments period.

As far as needing financial help after the payments had stopped, of the 179 families who answered that they had needed help 94 did not apply for help; only 8 applied for help and did not get it. Finally, families were asked to make a before-and-after comparison. About 88 percent estimated that they were as well off or better off after the experiment. Although it may be that the simple passage of time improved the experimental families' absolute level of well-being, these figures provide one more general indication that short-term experiments can be conducted without serious adverse effects on participants.

3. RESULTS FROM THE NEW JERSEY EXPERIMENT

This section on results will first show the movement of the actual negative income tax payments made to the families over time. Certain descriptive statistics showing the characteristics of the sample and changes in these characteristics over time will follow. Third, the labor-supply response by family members will be described. Then a summary of the results will be presented.

Movement of the Payments Over Time. Table 5 shows the average payment levels over the period of the experiment by site, by ethnic group, and (for the second year) by experimental plan. These payments show a mildly rising trend. When it is remembered, however, that a cost of living correction was made to the guarantee levels every year, amounting to 5.5 percent in September 1969, 5.9 percent in October, 1970, and 4.1 percent in September, 1971, and that, further, the experimental period was a period of rising unemployment, the small extent of the increase in payments is evidence that there was no widespread decrease in work effort, nor substantial falsification of income reports.

TABLE 5

AVERAGE PAYMENTS PER FOUR-WEEK PERIOD,

CONTINUOUS HUSBAND-WIFE FAMILIES BY SITE (DOLLARS)

	All Sites	Trenton	Paterson-Passaic	Jersey City	Scranton
First year	91.03	69.93	79.43	107.80	91.46
Second year	93.25	71.91	80.67	109.86	94.72
Third year	96.84	58.67	84.92	120.35	98.26
Percentage change, first to third year	6.4	-16.1	6.9	11.6	5.2

AVERAGE PAYMENTS PER FOUR-WEEK PERIOD,

CONTINUOUS HUSBAND-WIFE FAMILIES BY ETHNIC GROUP (DOLLARS)

	All	White	Black	Spanish-speaking-
First year	91.03	87.65	97.65	86.96
Second year	93.25	91.03	96.59	92.23
Third year	96.84	90.11	102.83	100.32
Percentage increase, first to third year	6.4	2.8	5.3	15.4

TABLE 5 (Cont'd.)

AVERAGE PAYMENTS IN DOLLARS PER FOUR-WEEK PERIOD,

CONTINUOUS HUSBAND-WIFE FAMILIES BY PLAN

(SECOND EXPERIMENTAL YEAR)

	Tax Rate (percent)		
	30	50	70
Guarantee level (Percent)			
125	no plan	187.28	no plan
100	no plan	123.72	66.07
75	103.54	44.17	34.91
50	46.23	21.66	no plan

Source: "An Overview of the Labor Supply Results," by Albert Rees, in Volume I of The Final Report of the New Jersey Graduated Work Incentive Experiment.

<u>Descriptive Statistics of the Sample</u>. Table 6 displays the distribution of the total sample by ethnic group, by site, and by income stratum. It reflects the facts mentioned earlier that virtually all the Scranton sample is white and that Paterson-Passaic contains a heavy concentration of Spanish-speakers. It also shows that less than a third of the sample was initially below the poverty line in terms of their income. Most of the analysis in the Final Report was done using a subsample of continuous husband-wife families. Table 7 shows similar distribution characteristics for that subsample of 693 families. This group of households contained the same husband-wife combination throughout the experiment and remained active questionnaire respondents through the whole three years.

Tables 8-14 provide descriptive statistics on the basic income and labor supply characteristics of the sample. Very general indications of the nature of the response to the experiment can be drawn from these tables. It should be noted that the divergence between controls and experimental averages often comes as a result of movement of both groups in the same direction but at different rates. Clearly there is no gross evidence of widespread or major abandonment of work on the part of any of the experimental groups. There is, however, a consistent pattern of slightly reduced labor supply.

Family characteristics are displayed in Tables 8-12; total income, total earnings, total hours, family size and number of employed persons. Average values are tabulated for

TABLE 6

THE NEW JERSEY-PENNSYLVANIA SAMPLE CHARACTERISTICS:

TOTAL SAMPLE

(percents in parentheses)

	All	White	Black	Spanish-speaking
TOTAL	1357	440	502	415
Negative Income Tax Plan				
50-30	48(3.5)	19(4.3)	19(3.8)	10(2.4)
50-50	73(5.4)	15(3.4)	28(5.6)	30(7.2)
75-30	101(7.4)	26(5.9)	41(8.1)	34(8.2)
75-50	117(8.6)	33(7.5)	43(8.6)	41(9.9)
75-70	85(6.3)	31(7.0)	38(7.6)	16(3.9)
100-50	77(5.7)	22(5.0)	32(6.4)	23(5.5)
100-70	86(6.3)	25(5.7)	34(6.8)	27(6.5)
125-50	138(10.2)	61(13.9)	47(9.4)	30(7.2)
Original Controls	491(36.2)	196(44.5)	151(30.1)	144(34.7)
New Controls	141(10.4)	12(2.7)	69(13.7)	60(14.5)
Site				
Trenton	159(11.7)	25(5.7)	105(20.9)	29(7.0)
Paterson-Passaic	490(36.1)	49(11.1)	194(38.6)	247(59.5)
Jersey City	390(28.7)	52(11.8)	199(39.6)	139(33.5)
Scranton	318(23.4)	314(71.4)	4(0.8)	0(0)
Income Stratum				
I (0-99 percent of poverty)	414(30.5)	119(27.0)	139(27.7)	156(37.6)
II (100-124 percent of poverty)	454(33.5)	153(34.8)	173(34.5)	128(30.8)
III (125-150 percent of poverty)	489(36.0)	168(38.2)	190(37.8)	131(31.6)

Source: Final Report of the New Jersey Graduated Work Incentive Experiment, by Harold Watts et al., Institute for Research on Poverty, 1974, Volume I, Part B, Chapter I.

TABLE 7

THE NEW JERSEY-PENNSYLVANIA SAMPLE CHARACTERISTICS:

CONTINUOUS HUSBAND-WIFE SAMPLE

	All	White	Black	Spanish-speaking
	(percent of relevant total in parentheses)			
TOTAL	693	310	234	149
Negative Income Tax Plan				
50-30	27(3.9)	13(4.2)	8(3.4)	6(4.0)
50-50	32(4.6)	11(3.5)	12(5.1)	9(6.0)
75-30	60(8.7)	22(7.1)	23(9.8)	15(10.1)
75-50	65(9.4)	24(7.7)	25(10.7)	16(10.7)
75-70	48(6.9)	24(7.7)	21(9.0)	3(2.0)
100-50	44(6.3)	20(6.5)	14(6.0)	10(6.7)
100-70	53(7.6)	21(6.8)	17(7.3)	15(10.1)
125-50	96(13.9)	46(14.8)	31(13.2)	19(12.8)
Controls	268(38.7)	129(41.6)	83(35.5)	56(37.6)
Site				
Trenton	60(8.7)	12(3.9)	38(16.2)	10(6.7)
Paterson-Passaic	158(22.8)	30(9.7)	59(25.2)	69(46.3)
Jersey City	236(34.0)	32(10.3)	134(57.3)	70(47.0)
Scranton	239(34.5)	236(76.1)	3(1.3)	0(0)
Income Stratum				
I (0-100 percent of poverty)	179(25.8)	71(22.9)	53(22.6)	55(36.9)
II (101-125 percent of poverty)	237(34.2)	105(33.9)	85(36.3)	47(31.5)
III (126-150 percent of poverty)	277(40.0)	134(43.2)	96(41.0)	47(31.5)

Source: Same as for Table 6.

TABLE 8

AVERAGE TOTAL FAMILY INCOME EXCLUSIVE OF EXPERIMENTAL PAYMENTS OR WELFARE ($ PER WEEK)

	Total Sample at Pre-enrollment (n = 1213)	Continuous Husband-Wife Sample (n = 693)			
		Pre	1st Year	2nd Year	3rd Year
Total Sample	98.27	103.09	117.31	130.71	143.63
Experimentals	98.20	102.30	117.32	129.45	140.28
Controls	98.38	104.34	117.30	132.71	148.93
Whites					
Experimentals	109.59	111.09	118.93	133.18	145.84
Controls	104.45	111.51	123.92	144.64	165.12
Blacks					
Experimentals	93.50	99.21	121.41	133.39	144.07
Controls	93.59	97.00	113.13	121.82	133.54
Spanish-speaking					
Experimentals	91.89	90.18	107.56	115.78	123.31
Controls	95.14	98.74	108.25	121.35	134.44
By Site					
Trenton	87.90	91.45	107.46	123.32	142.62
Paterson-Passaic	91.78	96.16	118.74	130.57	139.55
Jersey City	99.52	102.44	119.15	126.37	137.86
Scranton	108.60	111.22	117.03	136.94	152.26

Source: Same as for Table 6.

TABLE 9

AVERAGE FAMILY EARNINGS ($ PER WEEK)

	Total Sample at Pre-enrollment (n = 1213)	Continuous Husband-Wife Sample (n = 693)			
		Pre	1st Year	2nd Year	3rd Year
Total Sample	89.00	94.93	107.06	113.55	125.43
Experimentals	88.86	95.18	108.16	113.63	123.93
Controls	89.20	94.54	105.31	113.41	127.80
Whites					
Experimentals	96.18	100.24	106.67	114.49	126.26
Controls	92.11	98.02	107.37	122.17	143.38
Blacks					
Experimentals	86.71	94.40	115.30	120.24	128.38
Controls	86.11	89.73	104.87	104.26	109.83
Spanish-speaking					
Experimentals	83.65	86.60	99.47	101.23	112.19
Controls	88.49	93.64	101.23	106.80	118.54
By Site					
Trenton	87.90	91.45	102.46	107.55	116.30
Paterson-Passaic	81.37	86.90	107.64	109.68	123.01
Jersey City	93.23	98.45	113.35	115.92	123.55
Scranton	93.37	97.64	101.62	115.27	131.17

Source: Same as for Table 6.

TABLE 10

AVERAGE TOTAL FAMILY HOURS WORKED PER WEEK

	Total Sample at Pre-enrollment (n = 1213)	Continuous Husband-Wife Sample (n = 693)			
		Pre	1st Year	2nd Year	3rd Year
Total Sample	39.54	40.6	40.9	40.4	42.4
Experimentals	39.26	40.6	39.9	39.2	50.8
Controls	39.95	40.7	42.6	42.4	44.9
Whites					
Experimentals	41.49	42.3	39.6	39.8	42.2
Controls	40.66	42.2	44.0	45.5	50.9
Blacks					
Experimentals	37.94	40.3	41.8	39.9	40.5
Controls	37.13	36.2	40.6	38.7	37.3
Spanish-speaking					
Experimentals	38.56	37.8	37.2	36.8	38.4
Controls	41.95	43.9	42.2	40.7	42.3
By Site					
Trenton	41.79	44.3	42.8	39.8	39.8
Paterson-Passaic	36.38	36.5	38.4	38.8	40.9
Jersey City	40.01	40.1	41.6	39.5	39.7
Scranton	41.85	43.0	41.4	42.6	46.6

Source: Same as for Table 6.

TABLE 11

AVERAGE FAMILY SIZE

	Total Sample at Pre-enrollment (n = 1213)	Continuous Husband-Wife Sample (n = 693)			
		Pre	1st Year	2nd Year	3rd Year
Total Sample	5.89	6.14	6.19	6.25	6.3
Experimentals	6.02	6.27	6.31	6.36	6.42
Controls	5.70	5.93	6.01	6.06	6.13
Whites					
Experimentals	5.69	5.86	5.86	5.87	5.87
Controls	5.44	5.71	5.79	5.81	5.80
Blacks					
Experimentals	6.52	7.05	7.08	7.18	7.31
Controls	6.02	6.29	6.43	6.47	6.51
Spanish-speaking					
Experimentals	5.73	5.82	5.92	5.99	6.03
Controls	5.72	5.88	5.89	6.04	6.31
By Site					
Trenton	6.13	6.38	6.47	6.63	6.79
Paterson-Passaic	5.74	6.09	6.17	6.19	6.22
Jersey City	6.28	6.62	6.64	6.72	6.85
Scranton	5.51	5.64	5.70	5.72	5.70

Source: Same as for Table 6.

TABLE 12

NUMBER OF EMPLOYED PERSONS PER FAMILY

	Total Sample at Pre-enrollment (n = 1213)	Continuous Husband-Wife Sample (n = 693)			
		Pre	1st Year	2nd Year	3rd Year
Total Sample	1.075	1.059	1.102	1.080	1.115
Experimentals	1.084	1.071	1.066	1.035	1.067
Controls	1.061	1.041	1.159	1.153	1.190
Whites					
Experimentals	1.142	1.105	1.070	1.075	1.135
Controls	1.066	1.047	1.171	1.258	1.374
Blacks					
Experimentals	1.061	1.053	1.103	1.045	1.043
Controls	1.040	.976	1.163	1.069	.982
Spanish-speaking					
Experimentals	1.052	1.032	.997	.941	.976
Controls	1.076	1.125	.955	1.036	1.076
By Site					
Trenton	1.134				
Paterson-Passaic	1.018		not available		
Jersey City	1.054				
Scranton	1.145				

Source: Same as for Table 6.

TABLE 13

AVERAGE NUMBER OF EMPLOYED MALE HEADS PER FAMILY

(MAXIMUM = 1)

	Total Male Head Sample at Pre-enrollment (n = 1160)	Continuous Husband-Wife Sample (n = 693)			
		Pre	1st Year	2nd Year	3rd Year
Total Sample	.873	.885	.890	.868	.841
Experimentals	.866	.887	.895	.863	.836
Controls	.885	.881	.882	.876	.849
Whites					
Experimentals	.881	.906	.884	.856	.826
Controls	.895	.907	.884	.886	.876
Blacks					
Experimentals	.870	.861	.911	.881	.829
Controls	.825	.783	.828	.840	.786
Spanish-speaking					
Experimentals	.880	.892	.890	.847	.866
Controls	.929	.964	.955	.906	.906
By Site					
Trenton	.831	.817	.883	.825	.771
Paterson-Passaic	.836	.816	.883	.831	.834
Jersey City	.908	.919	.905	.909	.858
Scranton	.890	.912	.881	.863	.846

Source: Same as for Table 6.

TABLE 14

AVERAGE HOURS WORKED BY MALE HEAD PER WEEK

	Total Male Head Sample at Pre-enrollment (n = 1160)	Continous Husband-wife Samples (n = 693)			
		Pre	3rd Year	2nd Year	3rd Year
Total Sample	33.53	34.66	34.18	33.05	33.04
Experimentals	32.92	34.56	34.19	32.81	32.61
Controls	34.43	34.82	34.15	33.43	33.73
Whites					
Experimentals	34.09	35.33	33.92	32.44	32.10
Controls	35.54	36.87	35.33	34.67	36.22
Blacks					
Experimentals	31.80	34.11	35.27	33.17	32.10
Controls	30.58	29.59	30.66	30.17	28.86
Spanish-speaking					
Experimentals	33.04	33.83	32.97	32.94	34.44
Controls	36.68	37.84	36.58	35.42	35.20
By Site					
Trenton	33.02	34.53	34.63	31.76	28.93
Paterson-Passaic	30.62	30.66	31.66	31.21	32.50
Jersey City	35.65	35.72	35.21	34.20	33.26
Scranton	34.46	36.30	34.70	33.46	34.22

Source: Same as for Table 6.

all controls and experimental families and also for the respective ethnic groups. The averages for the four sites are also shown (except for Table 12). For the continuous husband-wife sample the averages are shown as of the pre-enrollment survey and for the three separate years of the experimental period. For comparison the pre-enrollment averages are shown for all initially-enrolled families (excluding new controls and three families whose pre-enrollment questionnaire was lost). The continuous sample is shown to have somewhat larger families with higher income and earnings initially, but the differences are not great--around 5 percent.

The labor-supply status of the male head or husband in the continuous sample is exhibited in Tables 13 and 14. (The more inclusive sample of all male family heads at pre-enrollment is included for comparison.) Again, the differences are minor but uniformly in the direction of greater supply for the continuous sample.

The Technical Papers in the Final Report of the experiment provide extensive, rigorous, and sophisticated statistical analyses of the labor-supply results that we shall not attempt to summarize here. Instead we shall include tables showing treatment-control differentials as estimated by regression analysis for the Summary Report on the experiment released by HEW, along with parts of the commentary on the tables also contained in that report. These regressions include as control variables age, education, number of adults, number and ages of children, sites, and pre-experiment family earnings and labor supply.

Husbands' Labor Supply. Table 15 shows treatment-control differentials of married male heads of households for four measures of labor-supply response--labor force participation, employment, hours, and earnings--for the middle eight quarters of the experiment. The striking features of these results are that all the differentials are small in both absolute and relative terms--none exceed 10 percent of the control mean and most are less than five percent--and all are statistically insignificant (i.e., one cannot rule out the possibility that these differentials occurred purely by chance). There are no findings here to indicate a significant reduction in labor supply resulting from the experimental payments. Moreover, many of the differentials (including all of those for blacks) are positive, indicating greater labor supply among husbands in the treatment group than in the control group. It is also worth noting that the means for both groups indicate that the vast majority (approximately 95 percent) of the husbands were labor force participants, working, when employed, close to full time (37 to 40 hours per week).

The further statistical refinements on the data for husbands pursued in the Technical Papers still found no significant treatment effect for whites and blacks. They did, however, uncover a small but statistically significant decrease in labor force participation on the part of Spanish-

TABLE 15

HUSBAND TOTALS: REGRESSION

ESTIMATES OF DIFFERENTIALS IN LABOR FORCE PARTICIPATION, EMPLOYMENT, HOURS, AND EARNINGS FOR QUARTERS 3 to 10[1]

	Labor Force Participation Rate[2]	Employment Rate	Hours Worked per Week	Earnings per Week
White				
Control group mean	94.3	87.8	34.8	100.4
Absolute differential	-.3	-2.3	-1.9	.1
Treatment group mean	94.0	85.5	32.9	100.5
Percent differential	-.3	-2.6	-5.6	.1
Black				
Control group mean	95.6	85.6	31.9	93.4
Absolute differential	0	.8	.7	8.7
Treatment group mean	95.6	86.4	32.6	102.1
Percent differential	0	.9	2.3	9.3
Spanish-speaking				
Control group mean	95.2	89.5	34.3	92.2
Absolute differential	1.6	-2.4	-.2	5.9
Treatment group mean	96.8	87.1	34.1	98.1
Percent differential	1.6	-2.7	-.7	6.4

[1]The data for this table consist of 693 husband-wife families who reported for at least 8 of the 13 quarters when interviews were obtained. Percent differentials are computed using the mean of the control group as base.

[2]This includes those employed and unemployed. Someone is unemployed if he is actively seeking employment, waiting recall from layoff or waiting to report to a new wage or salary job.

Note: For a complete explanation of the data and methods used for these calculations, please refer to the HEW Summary Report (December, 1973).

speaking husbands. Significant treatment effects were again found for Spanish-speaking husbands for hours worked per week. If one evaluates the estimated response function[1] for an average Spanish-speaking husband on a plan with a basic benefit equal to the poverty line and a 50 percent implicit tax rate, the treatment effect on weekly hours worked is a reduction of 3.2 hours (mean hours worked by Spanish-speaking control husbands were 34.3). A similar calculation for white husbands yields a statistically significant reduction of 2.4 hours per week. For black husbands there was once again no significant treatment effect.

Much of the reduction in hours among Spanish-speaking husbands can be accounted for by declines in their employment rate (that is, the fraction of all Spanish-speaking husbands in the experimental population who were employed). This implies that Spanish-speaking husbands were unemployed more when in the treatment group, a result which is given independent confirmation when data on unemployment are analyzed directly. For white husbands, whose hours were reduced as

---

[1]The "response functions" on which the results presented in this section are based are regression equations relating the labor supply response variables to a set of control variables and the basic benefit levels and implicit tax rates of the experimental plans. These regressions were estimated using data from all continuous husband-wife families, in all plans and the control group. By inserting specific values of the control and treatment variables in these equations, one can predict the labor supply response of a particular type of family on a particular plan. References to responses under a specific plan are based on this type of calculation. In general, these predictions will be more precise than those based only on data from families in a particular plan.

noted above, the employment effect was small (and positive) so that all the experimental effect would appear to be in hours worked per week for those at work. As yet we do not know if this result arises from less overtime work, a reduction in multiple job holding, or some other source.

Viewing the results by experimental plan, it was found that the reduced labor supply for Spanish-speaking husbands varied, as we would expect, with the implicit tax rate--higher implicit tax rates produced substantially. stronger disincentives. For whites the reverse was true--the largest disincentives were estimated for plans with the lowest implicit tax rates. In neither case was there a strong or consistent ordering by basic benefit level; indeed, the most generous plan (125-50) showed the smallest treatment effects. Overall, then, the experiment produced no consistently significant effects by implicit tax rate or basic benefit. These results do not, of course, allow prediction of the labor supply effects of implicit tax rates or basic benefits outside the range employed in the experiment--that is, implicit tax rates below 30 percent or above 70 percent, or basic benefits less than 50 percent, or greater than 125 percent of the poverty line.[1]

By far the most surprising result of the analysis for husbands is the complete failure to find any significant effect for blacks, despite the fact that black husband-wife

---

[1]Even results for the 70 percent tax rate must be interpreted skeptically, because (as mentioned above) of the very small numbers of families in the 70 percent tax rate cells who were below their breakeven point and not receiving welfare.

families received slightly larger average payments than the other two ethnic groups. Indeed, the estimated supply response for blacks is not only insignificant, but preponderantly positive. The data indicate that earnings of the black control group increased more slowly over the course of the experiment than those of the other control and treatment groups. Thus, when treatment-control comparisons are made for blacks the differential in favor of the treatment group is noticeably large. This kind of finding for blacks is not limited to husbands; it recurs in the analysis of other components of the household. We have no plausible explanation for this outcome.

The Labor-Supply Response of Wives. Table 16 shows the regression results for wives--predominantly negative labor supply differentials. These were small in absolute magnitude, but, because of the low levels of market supply of wives, these differentials represent relatively large percentage differentials--at least for white and Spanish-speaking wives.[1] Even so, only two of the differentials shown in the Table--those for labor force participation and employment rates of white wives--are statistically significant. This lack of significance reflects the small absolute size of the differentials and the small sample sizes of working wives in each of the three ethnic groups; for example, in any given survey

---

[1]The means presented in the tables are averages over all individuals within a given group, including non-workers. Corresponding means for workers only can be readily calculated from the numbers presented. For example, while all white wives worked an average of 4.5 hours per week, the 17.1 percent of the control group who were employed worked an average of 26.3 (4.5/.171) hours per week.

TABLE 16

WIFE TOTALS: REGRESSION

ESTIMATES OF DIFFERENTIALS IN LABOR FORCE PARTICIPATION, EMPLOYMENT, HOURS, AND EARNINGS FOR QUARTERS 3 to 10[1]

	Labor Force Participation Rate[2]	Employment Rate	Hours Worked per Week	Earnings per Week
White				
Control group mean	20.1	17.1	4.5	9.3
Absolute differential	-6.7*	-5.9*	-1.4	-3.1
Treatment group mean	13.4	11.2	3.1	6.2
Percent differential	-33.2	-34.7	-30.6	-33.2
Black				
Control group mean	21.1	16.8	5.0	10.6
Absolute differential	-.8	-.3	-.1	.8
Treatment group mean	20.3	16.5	4.9	11.4
Percent differential	-3.6	-1.5	-2.2	7.8
Spanish-speaking				
Control group mean	11.8	10.7	3.4	7.4
Absolute differential	-3.8	-5.2	-1.9	-4.1
Treatment group mean	8.0	5.5	1.5	3.3
Percent differential	-31.8	-48.3	-55.4	-54.7

[1, 2] See notes to Table 15.

*Significant at the .95 level (two-tailed test).

week there were only about 15 working wives among the Spanish-speaking families in the entire sample. It is important to note that the labor supply of wives in the experiment as reflected by both of these measures, particularly labor force participation, are well below their average values for the population as a whole. For example, the pre-enrollment labor force participation rates of 16.0 percent and 13.4 percent for treatment and control wives, respectively, are less than one-half their values for all married women in the population. This results from the way in which the sample was selected. Only families with income less than one and one-half times the poverty line were admitted to the sample. Therefore, families with multiple earners had a low probability of selection. In addition, because the poverty line is adjusted upward as family size increases, the higher-income families in the experiment were likely to have larger families and younger children. Both these factors lead to an underrepresentation of working wives. Because pre-enrollment labor supply was quite small the absolute differentials seem large indeed in percentage terms.

In distinguishing among experimental plans, as was done in the Technical Papers, responses were generally consistent with expectations. For all wives the estimated negative response is consistently larger the more generous the plan, and the differences in response by plan are usually significant. A similar comparison by implicit tax rates found larger effects the higher the implicit tax rate, but these differences were usually small and never significant.

The estimated effects on labor supply of wives are subject to two rather different interpretations. The average estimated reduction in labor-force participation for all wives referred to above is 3 percentage points; for white wives it is 8 percentage points. These do not represent large absolute changes taken alone. But, because the mean participation rate for all control wives is only 17 percent, the estimated percentage reduction in labor supply for all wives in the treatment group (compared to controls) is 20 percent, and, for white wives, it is a sizeable 50 percent.

It should be noted that these estimated effects may be larger than those to be expected in an otherwise similar but permanent income maintenance program. For the control families, no more than 19 percent of wives were in the labor force in any one quarter, but 41 percent were in the labor force in at least one of the 13 quarters (counting pre-enrollment). In other words, this is a group that enters and leaves the labor force frequently. The experimental treatment creates a strong incentive to concentrate periods out of the labor force during the life of the experiment. A permanent program might therefore be expected to have a somewhat smaller impact.

The Family. Table 17 shows similar mean labor-supply differentials for the family as a whole--preponderantly negative but again relatively small.[1] In no case do the differentials exceed 14 percent of the control mean, and most are less than 10 percent. All the differentials for white families except for the earnings measure are statistically significant, while none of those for black or Spanish-speaking families are significant.

Earnings is particularly important as a labor supply measure for the family in that it provides a natural way to value or weight the hours worked by different family members; the weight is the wage rate of each member. Unfortunately, there is a possible bias in the experiment's measurement of the earnings variable not present in the other measures. Treatment families filled out an income report form every four weeks, while control families did not. The treatment families may therefore have learned more quickly than control families that what was to be furnished was gross rather than net earnings (that is, earnings before taxes and other deductions, <u>not</u> take-home pay). If this were the case, earnings in the treatment group (since gross exceeds net) would appear greater, relative to control earnings, than they actually are. This differential learning process could have caused a spurious differential in earnings in favor of the treatment group, especially during the early

---

[1]Family means and differentials include the labor supply of all workers in the family, not just husband and wife.

TABLE 17

FAMILY TOTALS: REGRESSION

ESTIMATES OF DIFFERENTIALS IN LABOR FORCE PARTICIPATION, EMPLOYMENT, HOURS, AND EARNINGS FOR QUARTERS 3 to 10[1]

	Number in Labor Force per Family[2]	Number Employed per Family	Hours Worked per Week	Earnings per Week	Percent of Adults in the Labor Force, per Family	Percent of Adults Employed, per Family
White						
Control group mean	1.49	1.30	46.2	124.0	57.6	51.1
Absolute differential	-.15*	-.18*	-6.2*	-10.1	-5.3*	-6.1*
Treatment group mean	1.34	1.12	40.0	113.9	52.3	45.0
Percent differential	-9.8	-13.9	-13.4	-8.1	-9.1	-12.0
Black						
Control group mean	1.38	1.17	41.7	114.0	54.3	46.9
Absolute differential	-.07	-.07	-2.2	4.1	-1.6	-1.6
Treatment group mean	1.31	1.10	39.5	118.1	52.7	45.3
Percent differential	-5.4	-6.1	-5.2	3.6	-2.9	-3.3
Spanish-speaking						
Control group mean	1.15	1.04	39.0	102.4	48.9	44.7
Absolute differential	.08	-.02	-.4	5.0	2.4	-1.0
Treatment group mean	1.23	1.02	38.6	107.4	51.3	43.7
Percent differential	6.7	-1.5	-.9	4.9	5.0	-2.2

[1, 2] See notes to Table 15.

*Significant at the .99 level (two-tailed test).

part of the experiment. Therefore, the results for hours worked and labor-force participation may be more reliable than for earnings.

In the more sophisticated analysis in the Technical Papers, hours worked and earnings both showed a significant reduction for white families, ranging from 8 to 16 percent for hours and 8 to 12 percent for earnings. For blacks, the earnings effects were significantly positive, rising by 9 to 13 percent. Effects on hours worked by black families are small and show no consistent pattern; in one analysis a decline of 3 percent was found, while in another an increase of 1 percent appeared.

For Spanish-speaking families estimates of significant hours reductions in the neighborhood of 2 percent to 6 percent were found, while earnings were estimated to fall anywhere from 2 percent to 28 percent. These estimates are based on evaluation of the estimated response functions for families in plans with a 50 percent implicit tax rate.

In parts of the analysis the statistically predicted variance of family income was included as a control variable. This variable represents the fluctuation in income over time--for example, from $200 per month in February to $600 per month in July for a construction laborer. Such a variable was included for two reasons. First, families with variable income may have weaker attachments to the labor force, and therefore the experimental payments may have a stronger effect on their behavior. Second, variation in income gives the family experience with the effect of the implicit tax rate on the level of payments. This variance of income measure had a highly significant effect on the labor supply of whites. The more variable was income, the more labor supply declined. Other ethnic groups did not evidence such behavior.

The results for white families are thus consistent with those from the separate analyses of husbands and wives in that significant negative effects on labor supply are found. For blacks, the results again show predominantly positive responses, though not consistently so for hours worked. For Spanish-speaking families, the labor supply effects are negative, though generally smaller and less significant than for whites.

Summary of Findings. There was no widespread withdrawal from work on the part of the experimental group. This is clear from the fact that average benefit payments to the experimental families increased over the period of the experiment by less than the cost-of-living correction built into the benefit calculation. In the first year of the experiment, the average four-weekly payment was $92. In the third year this had increased by only 3.8 percent, to $96.

The most important group for any national income maintenance policy with respect to the potential cost of such a program is that constituted by the non-aged able-bodied males with family responsibilities. These are the people with the most solid attachment to the labor force. These are the people with the most labor to withdraw. These are the people about whom there is the most widespread fear that, given an income alternative, they will decide not to work. As it turned out, the effect for this group was almost undetectable. Over the central two years of the experiment (the period least contaminated by start-up and end effects), the employment rate for male family heads in the experimental group was only 1.5 percent less than that for the controls. For the number of hours worked per week the difference amounted to just over 2 percent. For earnings per week the experimentals actually were higher by 6.5 percent. This finding is at least partly spurious, due to a probable accelerated learning effect whereby experimentals learned to report gross rather than net earnings faster than controls. It also appears partly due to the fact that the younger and better-educated experimentals were able to use the insurance provided by the payments to look for (and find) better, more stable, jobs.

The second group in terms of policy interest is the wives. The average family size in the sample was six, so these wives must be considered on the average to be mothers of four children. These wives had lower labor force participation than the national average, about 15 percent working at any survey point. For this group the differential between experimentals and controls was substantial, with experimental wives working 23 percent fewer hours per week than controls, their employment rate being 24 percent less, and their average earnings per week 20.3 percent less. This should not be regarded necessarily as an adverse outcome, given the fact that wives in six-person families work very hard inside the home, and that this work could well be more beneficial (cost-effective) from a national point of view than low-wage market labor. It should be noted, in addition, that although this relative reduction is large, it in fact starts from an average figure of only 4.4 hours a week. So from the point of view of family labor supply and national costs, it is not a large absolute change.

This brings us to total family labor supply--a composite of market work by the husband, the wife, and other adult family members. Predictably, these estimates lie between those for husbands and wives. Over the central two years, the number employed per family was 9.5 percent less for experimental families than controls. The hours worked per week per family were 8.7 percent less for experimentals than controls. The average earnings per week were almost the same. This disincentive was almost entirely made up of relative work withdrawal by secondary earners--wives who decided to work more inside the home, teenagers who may have been enabled by the payments to stay in school longer, and older workers who were able to take it a bit easier. As such, the disincentive effect may well be considered to be socially useful.

The analysis has shown a persistent difference in response according to ethnic groups--white, Black, and Spanish-speaking. Such disincentive as was found for husbands was restricted mainly to whites. The substantial disincentive for wives was also largely due to white wives. For both males and females the Spanish-speaking showed more disincentive than the Blacks, who showed none. No satisfactory explanation has yet been found for this difference. It is apparent that Black controls had an unusually bad labor-market experience in the last year of the experiment, both compared with Black experimentals and with the controls from the other two ethnic groups. Further research is underway to try and pin down the causes for this ethnic difference.

Response in areas other than labor force participation were generally slight. In the area of expenditures, the experimentals showed a tendency to move from public to private rental housing, and to buy relatively more homes. They also bought somewhat more furniture and other durables, and consequently incurred more debt.

In the area of psychological and sociological responses, the effects were negligible. Cash assistance at the levels involved in this study does not appear to have a systematic effect on the recipients' health, self-esteem, social integration, or perceived quality of life, among many other variables. Nor does it appear to adversely affect family composition, marital stability, or fertility rates.

What we can say with certainty is that these benefits represented a net increase in family income, allowing these families greater command over material goods and services, and enhancing their economic well-being. The anti-poverty effectiveness of the payments was not seriously vitiated by offsetting reductions in earnings due to reduced work effort.