INSTITUTE FOR RESEARCH ON POVERTY DISCUSSION PAPERS

196-74

BASIC LABOR SUPPLY RESPONSE FINDINGS FROM THE URBAN EXPERIMENT (New Jersey-Pennsylvania)

Harold W. Watts

Glen G. Cain

UNIVERSITY OF WISCONSIN - MADISON

BASIC LABOR SUPPLY RESPONSE FINDINGS FROM THE URBAN EXPERIMENT (New Jersey-Pennsylvania)

Harold W. Watts

Glen G. Cain

February 1974

This report of research was presented at the annual meetings of the American Economic Association (joint session with the Econometric Society). New York, December 30, 1973

Author's Note: This is a composite paper, incorporating pieces from a variety of individual researchers working on the project, including, besides the authors, Robert Avery, Robinson Hollister, David Horner, Charles Mallar, Walt Nicholson, Dale Poirier, Albert Rees, Felicity Skidmore, Judith Wooldridge.

In addition to the above, we would like to note our thanks to other personnel at the Institute for Research on Poverty, Mathematica, and OEO, without whom neither the administration of the experiment nor the final report could have been successfully completed.

Any formal citations to results should be made to the full report as follows: Final Report of the Graduated Work Incentives Experiment, Institute for Research on Poverty, December 1973. A summary version of the labor supply results will be available as a Special Supplement to the Spring 1974 issue of the Journal of Human Resources. A special summary paper released on the experiment by the Department of Health, Education, and Welfare is also available, entitled Summary Report: The New Jersey Graduated Work Incentive Experiment, December, 1973. All inquiries concerning these publications should be directed to Ms. Angela Keynon, Institute for Research on Poverty, University of Wisconsin, Madison, Wisconsin 53706. The research reported here was supported by funds granted to the Institute for Research on Poverty at the University of Wisconsin by the Office of Economic Opportunity pursuant to the provisions of the Economic Opportunity Act of 1964.

The opinions expressed are those of the various researchers and do not necessarily reflect the opinions of the Institute for Research on Poverty or the Department of Health, Education, and Welfare.

ABSTRACT

The New Jersey Graduated Work Incentive Experiment was designed to measure the response of households including an able-bodied male between 18 and 58 years of age to a set of negative income tax plans. The experiment took 6 years to complete, involved 8 negative income tax plans, some 1,350 experimental and control families, 5 cities, 3 years of payments in each site, and a cost of almost $8 million.

This paper reviews the labor-supply results for the 693 husband-wife families who reported continuously throughout the experiment and did not change their marital status. Results are analyzed separately for married men, married women, and the family as a whole. The regression estimated mean differences in several measures of labor supply show the following.

Male heads. All the experimental-control differentials are small in both absolute and relative terms. None exceed 10 percent of the control mean, most are less than 5 percent, and all are insignificant. Many of the differentials, including all those for blacks, are actually positive. The means for both experimentals and controls also indicate that approximately 95 percent of the husbands were labor-force participants, working, when employed, between 37 and 40 hours a week.

Wives. Experimental wives show predominantly negative labor supply differentials. They are small in absolute magnitude but, because of the low levels of market supply of wives, represent relatively large percentage differentials--at least for white and Spanish-speaking wives. Even

so, only labor force participation and employment rates for white wives are statistically significant.

Family. Mean labor-supply differentials for the family as a whole are preponderantly negative, but again relatively small. In no case, do the differentials exceed 14 percent of the control mean, and most are less than 10. All the differentials are significant for white families except for the earnings measure, but none of those for Black or Spanish-speaking families.

More complex statistical models are also analyzed in the paper, but the basic shape of the results is not thereby changed.

Labor Supply Effects of the New Jersey-Pennsylvania Graduated Work Incentive Experiment

Harold W. Watts and Glen G. Cain

Background

The New Jersey-Pennsylvania Graduated Work Incentive Experiment was designed to measure the response of households including an able-bodied male between 18 and 58 years of age (not going to school full-time, institutionalized, or in the armed forces) to a set of negative income tax plans. This experiment, the first of its kind, began enrolling families in 1968, and ended field operation in 1972. The Final Report was submitted to the Department of Health, Education, and Welfare, in December 1973.

The project grew from a great many sources, and historians will draw different paths for the decision-making that eventually led to the funding and fielding of the first negative income tax experiment.

However, the path which has become most apparent to the staff of the experiment itself leads from a clear imperative in the era of the early War on Poverty to do something about the welfare "mess." A major reform of a negative tax variety as outlined by Milton Friedman and elaborated and expanded by James Tobin was developed and proposed--but was rejected for various reasons, one of the major ones being concern about the effect of unconditional income supplements upon the labor supply of able-bodied persons among the recipient group. Of particular concern was the response of the working poor and near poor, typically families with more than one adult, because the single-parent family was already eligible for AFDC benefits under terms that were decidedly discouraging to gainful employment.

The labor supply response to variations in income supplements and net wage rates was perceived as an issue which could be studied scientifically. But it was also one in which existing theory and sources of empirical evidence seemed unable to provide either intellectually or politically convincing answers. It was clear that an experiment in which such a scheme was implemented offered the possibility of providing more convincing evidence; and despite the novelty of the idea, it seemed increasingly feasible the more it was discussed.

The operation, which finally took more than six years to complete, has involved some 1,350 experimental and control families, four cities (Trenton, Paterson-Passaic, and Jersey City, New Jersey, and Scranton, Pennsylvania), and a total cost of $7.9 million of which one-third represented direct cash payments to sample families.

As is well known, a negative income tax plan guarantees a certain payment (the guarantee) in the event that the household receives no other income, and reduces this payment by a certain percentage (the tax rate) of every dollar the household earns up to the income level (break-even level) at which the payment is reduced to zero. In the New Jersey-Pennsylvania Graduated Work Incentive Experiment, four guarantee levels were tested (50, 75, 100, and 125 percent of poverty line) and three tax rates (30, 50, and 70 percent). These were combined into eight negative income tax plans in all (with guarantee level-tax rate combinations of 50-30, 50-50, 75-30, 75-50, 75-70, 100-50, 100-70, 125-50). To be eligible, families had to have an income for the year preceding the experiment of not more than 150 percent of the poverty line.

Originally enrolled in the experiment were 1,216 families: 725 in the experimental groups and 491 in the control groups. After the experiment was underway, 141 new controls were added in Trenton and Paterson-Passaic (October 1969). The total number of families, including the control families, was made up of the following percentages in three ethnic groups: white, 32 percent; Black, 37 percent; and Spanish-speaking, 31 percent. The families were enrolled sequentially in the following sites: Trenton, New Jersey (August 1968); Paterson-Passaic, New Jersey (January 1969); Jersey City, New Jersey (June 1969); and Scranton, Pennsylvania (September 1969). Payments were recalculated every four weeks, according to income received and family size. The income support was continued for three years in each site.

The data came primarily from interviews with each experimental and control family, conducted four times a year for the three-year period over

which payments were being made. A follow-up interview was administered to all participants in both the experimental and the control groups three months after the transfer payments had ceased. While this interview included questions on labor force effort, its main purpose was to investigate the families' understanding of the experiment, and their reactions to the transfer payments and the interviews.

A Brief Overview of the Findings

A succinct summary of our findings is presented in Tables 1-3, which show a regression-estimated mean difference in several measures of labor supply for a selected subset of the sample observations; namely, 693 husband-wife families who met the criteria for continuous reporting. This subset of families was selected because they are a relatively homogeneous group representing the modal family type among the working poor, for whom the analysis is not complicated by the problems of changes in family composition and missing data. (Other aspects of the sample selection, time period analyzed, and regression specification are discussed below.) Negative differentials, both absolute and percentage, indicate smaller labor supply on the part of the experimental families compared with control families. Within each table, results are reported separately for each ethnic group which, it should be noted, showed important differences in responses.

The most striking features of the results for husbands (shown in Table 1 are, first, that all the differentials are quite small in both absolute and relative terms--none exceed 10 percent of the control

Table 1

Husbands: Adjusted Mean Estimates Derived from Regression Estimates of Differentials in Labor Force Participation, Employment, Hours, and Earnings for Quarters 3 to 10[a]

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

[a]The data for these tables consist of 693 husband-wife families who reported for at least 8 of the 13 quarters when interviews were obtained. The reported differentials in each measure of labor supply is the experimental treatment group mean minus the control group mean, as measured in a regression equation in which the following variables were controlled: age of husband, education of husband, number of adults, number of children, sites, pre-experiment labor supply variables of the husband. These means and the associated control-experimental differentials may therefore be interpreted as applicable to control and experimental groups with identical composition in terms of these variables. Percent differentials are computed using the mean of the control as base.

Table 2

Wives: Adjusted Mean Estimates Derived from Regression Estimates of Differentials in Labor Force Participation, Employment, Hours, and Earnings for Quarters 3 to 10[a]

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

[a]The data for these tables consist of 693 husband-wife families who reported for at least 8 of the 13 quarters when interviews were obtained. The reported differentials in each measure of labor supply is the experimental treatment group mean minus the control group mean, as measured in a regression equation in which the following variables were controlled: age of wife, education of wife, number of adults, number and ages of children, sites, pre-experiment family earnings (other than wife's) and pre-experiment labor supply variables of the wife. These means and the associated control-experimental differentials may therefore be interpreted as applicable to control and experimental groups with indentical composition in terms of these variables. Percent differentials are computed using the mean of the control as base.

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

Table 3

Family Totals: Adjusted Mean Estimates Derived from Regression Estimates of Differentials in Labor Force Participation, Employment, Hours, and Earnings for Quarters 3 to 10[a]

	Number in labor force per family	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 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**
Experimental 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 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
Experimental 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 mean	1.15	1.04	39.0	102.4	48.9	44.7
Absolute differential	.08	-.02	-.4	5.0	2.4	-1.0
Experimental 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

[a]The data for these tables consist of 693 husband-wife families who reported for at least 8 of the 13 quarters when interviews were obtained. The reported differentials in each measure of labor supply is the experimental treatment group minus the control group mean, as measured in a regression equation in which the following variables were controlled: age of husband, education of husband, education of wife, number of adults, number and ages of children, sites, and pre-experiment labor supply variables for the husband and wife. These means and the associated control-experimental differentials may therefore be interpreted as applicable to control and experimental groups with identical composition in terms of these variables. Percent differentials are computed using the mean of the control as base.

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

mean and most are less than five percent--and all are insignificant. In other words, there are no findings to indicate a significant reduction in labor supply resulting from the plans. Moreover, many of the differentials, including all of those for Blacks, are positive--indicating greater labor supply among husbands in the experimental group than in the control group. Finally, it is worth noting that the means for both groups indicate that the vast majority (approximately 95 percent) of the husbands were labor force participants, working close to full-time when employed (37 to 40 hours per week).

Experimental wives, presented in Table 2, showed predominantly negative labor supply differentials. These were small in absolute magnitude, but because of the low levels of market labor supply of wives, these differentials represent relatively large percentage differentials--at least for white and Spanish-speaking wives.* 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 week there were only about 15 working wives among the Spanish-speaking families in the entire sample.

Mean labor supply differentials for the family as a whole, shown

---

*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 hours per week (= 4.5/.171).

in Table 3, were preponderantly negative, but again were relatively small.* 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.

In summary, these results present a picture of generally small absolute labor supply differentials between the experimental and control groups as a whole. Only among wives, whose mean labor supply is quite small to begin with, are the differentials large in relative terms. Although these summary results do not begin to exploit the richness of the data and the design of the experiment, they do fairly reflect the principal conclusions of our empirical work regarding labor supply. We cannot, of course, expect that these summary results will be acceptable without a great deal more detailed work dealing with more complicated models and, in particular, an intensive examination of the different negative income tax plans. We turn to these issues next, after first describing our general strategy for the statistical models used as the basis for our estimation procedures and tests of hypotheses.

The Models for Statistical Analysis: General Considerations

The models we use for statistical analyses are derived from the traditional literature on labor supply, but special features of the data

---

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

call for several modifications. The typical classroom presentation of the theoretical model of an individual's labor supply is anchored in the convention of comparative statics. When applied empirically, it is the average or expected labor supply of the sample--hours worked, for example--which is assumed to represent an equilibrium state. The observed number of hours worked is assumed to be a function of the person's normal wealth status, normal wage earnings capacity, and a variety of variables which are impounded in a ceteris paribus clause. Strictly speaking, these variables are, of course, unobservable, and various proxies for them are sought.

This simple model, however, which employs only income and wage rates as economic arguments, is instructive as a point of departure for analyzing the experimental data. The experiment induces changes in family income and in the net wage rates facing family members, and this fact, along with the random assignment of families to treatment and control groups, justifies ignoring other variables in the model for our pedagogic purposes here.

Let L_t be the labor supply of the individual during the time period, t, of interest, and Y be the equilibrium amount of household income if all family members worked the equilibrium amount of time (L* for the individual in question) and if nonlabor income were similarly at its normal level. Let W equal the market rate the individual can earn. Ignoring other variables and a stochastic error term, we have:

$$L_t = f(W,Y) \qquad (1)$$

The function is often expressed and estimated in a simple linear form:*

---

*The functional form could be derived if an operational function to describe the relation of goods and leisure to household utilities (or, alternatively, household production) were specified. One such utility function is used in this manner and described below, but in general we have adopted a more agnostic approach.

$$L_t = a_o + a_1 W + a_2 Y \qquad (2)$$

The variable, Y, is defined as:

$$Y = Y_o + WL^* \qquad (3)$$

where Y_o is the amount of family income from all other sources except the individual's earnings. (In the simplest case where there are no other family workers, Y_o would represent only nonlabor income. If other family members did work then this simple model could proceed without modification if we assumed that their wage rates and market labor supply decisions had no cross-price effects on the individual's labor market decisions. These are clearly abstractions from reality.)

A negative income tax plan will provide transfer payments, P, to families with incomes below the "breakeven level" (defined below), according to the following formula:

$$P = G - tY, \quad \text{for } Y < G/t = \text{breakeven} \qquad (4)$$

where G is the income guarantee at zero earned income and t is the plan's tax (or offset) rate on family income. When the transfer payments, P, are added to the family's income, the income and wage terms in the labor supply equation are changed as follows:

Substituting (3) and (4) into (2) and rearranging:

$$\begin{aligned} L_t &= a_o + (a_1 + a_2 L^*)(1-t)W + a_2[(1-t)Y_o + G] \qquad (5) \\ &= a_o + b_1(1-t)W + a_2[(1-t)Y_o + G] \\ &= a_o + b_1 W - b_1 tW + a_2 Y_o - a_2 tY_o + a_2 G \end{aligned}$$

Alternatively, we could write the equation in terms of t and P as:

$$L_t = a_o + b_1 W - b_1 tW + a_2 Y_o + a_2 P \qquad (6)$$

Since no allowance in the original model was made for differential responses to different <u>sources</u> of wage or income changes, it should not

be surprising that in (5) the coefficients of W and tW are the same except for sign and that the coefficients of Y_o, tY_o, and G are similarly equal in absolute value. In (6) the coefficients of Y_o and P are similarly restricted and furthermore, $a_2 = (b_1 - a_1)/L^*$. These restrictions are not in any serious sense imposed by theory, but are only reflections of our simplifying assumptions in applying one illustrative theoretical model. In any case, the restrictions can be relaxed and tested with the data. In the analyses of male heads-of-households discussed below some attention is given to (a) the construction of normal wage and income variables; (b) their effects on labor supply in comparison with the effects of the experimental variables, G, P, and t; and (c) their interactions with the experimental variables.

Another consequence of the simplifying assumptions made in this pedagogic example is that the effects of t and G are linear. In the experiment there are eight permissible combinations of the three tax rates and four guarantee levels (not including the control group), and we test for nonlinearities in the response to these programs. We also examine whether just being enrolled in a plan and being eligible to receive payments has any effect on labor supply behavior. Note that families with incomes above the breakeven level will not receive payments or face the experimental tax rates, but their work behavior may well be affected just by their being eligible for payments if their incomes were to decline.

Most of the statistical results we report are for a treatment specification that uses less than the eight variables which define each of the permissible plans. One can argue that the plans are sufficiently difficult to understand--particularly given changes in family size, fluctuations in

income above and below breakeven, the availability of welfare, and various "random shocks"--that accurate qualitative responses and approximately accurate quantitative measures are better achieved by a simpler functional form. A single anomalous cell value may be "corrected" by averaging it with better behaved adjacent cells. With a full specification of eight treatment variables, each plan is represented by a single cell value, independent of other cells. In a three-year experiment, however, it is perhaps unrealistic to expect to measure consistent and systematic responses to eight plans by a relatively small sample per plan, particularly when the responses differ by ethnic group.

Three other practical considerations may be mentioned briefly. First, we have not dealt with the distinction between a permanent nationwide income maintenance plan and a three-year experimental plan. Second, the parameters of state welfare plans have been ignored. Ways of dealing with both these problems are discussed later in this paper and in other papers in the full report. Third, traditional sources of data have not yielded satisfactory measures of normal wage or normal income, not to mention taste variables. Although the measures of these non-experimental variables are elusive for us as well, our repeated observations on income receipts and wage rates do offer some improvement regarding these important economic variables.

There are, of course, several easily measured variables, such as the individual's age, education, number of dependents, and others, which explain some of the variation in labor supply. However, a simple and more powerful device for controlling for the basic determinants of labor supply is to include the pre-experiment value of the labor supply variable

as an independent variable. The pre-experiment values of labor supply will partly reflect the effects of the available explanatory variables and in addition may represent other variables which are not satisfactorily measured, like normal income, normal wages, and taste variables. These considerations are represented in equation (7) which is the basic general model for the statistical analysis that follows:

$$L_{it} = f(X_{it}, Z_{it}, L_{i,t-1}, T_{it}, T_{it} Z_{it}) + \varepsilon_{it} \qquad (7)$$

where L = a measure of labor supply for the ith individual and experimental time period t;

X = a vector of "control" variables which are not presumed to interact with the treatment variables;

Z = a vector of "control" variables, including proxies for normal income and wages, which may both interact with the treatment and have separate effects;

L_{t-1} = a vector of pre-experiment values of the dependent variables;

T = a vector of treatment parameters;

ε = a residual term which is assumed to be uncorrelated with the deterministic part of the right-hand side of the model, (or, less restrictively, uncorrelated with T and TZ).

A final point about the specification of the statistical models concerns the advantage of the panel observations for thirteen quarters for measuring the labor supply response. We have the choice of examining the entire three-year period or selected subperiods. Subperiods can show the time pattern of responses. Also, a particular subperiod may be regarded as more representative of the operations of a legislated income maintenance plan. In this abbreviated version of the research we devote most attention to the middle 24 months of the experiment, quarters 3 to 10. This middle period is viewed as simulating a more "normal" course of a legislated plan because

the "start-up" and "wind-down" months of the experiment are ignored. The models are usually based on a single period of time for which we have a single observation per individual about his or her labor supply. The observation is some average measure over the time period.*

Male Heads

The labor supply or work effort response of male heads of families eligible for or receiving income subsidies such as a negative income tax is crucial from two points of view. First, the earnings of the male are typically the major source of earnings for poor and near-poor husband-wife families, and few such families have important amounts of income other than earnings. As the major earner, then, the husband has a large potential for labor force withdrawal in response to a transfer payment. In other words, a negative response large enough to negate the augmentation in money income from the transfer is possible for the primary earner, but less so for the secondary earners. Second, there is a popular view that any reduction in work-for-pay on the part of husbands with heavy family responsibilities is unrelieved either by the offsetting gains in output of work-at-home such as is expected from wives, or by investments toward future income such as is expected on the part of adult children (whose alternative may be successful completion of high school or added training of some sort). Whether these views are accurate is perhaps less important here than their prevalence. In any case, they place an importance on any weakening of the traditional "breadwinner" effort that has a different basis from concern about the work effort of secondary earners.

Married, non-aged males have typically been found to have high rates of labor-force participation and to be, moreover, quite insensitive to price

* An alternative specification which was also used extensively is based on a pooling of the quarterly observations. With 13 observations, one for each quarter, the potential number of observations for the full sample of 693 continuously reporting husband-wife families is, therefore, 9,009, although they are not, of course, "independent" observations.

(wage) and income variation. Hence, if the poor and near-poor male heads respond to such economic stimuli in about the same way as more general populations, and if the non-experimental studies have not somehow misled us about that response, we should not expect to find large responses for this group. With some complicating qualifications, this is about what is found in the analysis that follows. The over-all responses to tax rates that (on average) cut net wages in half and income guarantees that were equal to a substantial fraction of pre-experimental income are hardly detectable, and could be interpreted as so nonsignificant that further analysis is unwarranted. (See Table 1.)

Nevertheless, further analysis was undertaken which teased out and captured results consistent with and partly constrained by prior hypotheses about the nature of the supply response. The relative success of this effort suggests that there is some small average response, which is achieved by combining negligible responses around the breakeven level with more appreciable ones for husbands whose wages and potential earning power are well below the breakeven level.

It will be recalled from Table 1 in the introductory section that the basic contrasts for all experimental vs control families showed no significant differences for husbands. If one aggregates those estimates, using the sample frequencies of the ethnic groups for weights, the largest negative effect is in hours, and that is only -2.2 percent of the central group mean--well within the margin allowed for random sample variation. Both the size and the lack of significance of these differentials suggest there is relatively little grist to be ground in a more elaborate econometric mill. However, there are substantial differences in the size and terms

of the experimental treatments that are not distinguished in these averages. Also there has been no opportunity for the level of a family's earning capacity to affect the response to a given negative tax plan. It is certainly reasonable to expect any response to be substantially attenuated for a husband in a family whose income typically fluctuates around or above the breakeven level of its assigned treatment.

The remainder of this section will report on three more refined models which attempt to find significant and interpretable patterns which may have been obscurred in the cruder models. In the first of these the eight different guarantee/tax treatments are grouped according to their "generosity" over the income range represented in the sample, and separate response estimates are made for each. In the second a more elaborate model allowing for an interaction between earning capacity and the treatment is estimated. From this model it is possible to calculate estimates of the income and substitution effects of the negative tax treatment. Finally, the data are analyzed using a model that is severely constrained by a priori considerations derived from a model of utility maximization. Together the three approaches yield some additional insights into the form of any response to the experimental treatment. They do not, by relaxing or tightening the constraints on the estimated response, uncover evidence of sharp disincentives which would invalidate the general conclusion of small, almost undetectible responses for husbands with family responsibilities.

Experimental Effects by "Generosity of Plan"

Consider first separate responses for a tripartite division of experimental plans. The top three plans (125-50, 100-50, 75-30)

are denoted "High." The next three (100-70, 75-50, and 50-30) pay distinctly less over the relevant range and are called "Medium". The final two (75-70 and 50-50) pay almost nothing over most of the range and are called "Low". These two plans are moreover heavily dominated by public assistance support levels over most of the experimental period. Subsequently, the Low group will be deleted from the analysis of male head response on the ground that some 90 percent of experimental families on these two plans are either on welfare or have income above the break-even point. While their behavior may be of interest, it is not likely to tell us much about the response to experimental stimuli. Table 4 shows how the sample of continuous husband-wife families used here is distributed over this aggregation of experimental plans.

Table 4

Distribution of Sample Families by "Generosity" of Plan and Ethnicity

				Group Plan		
	All	Control	Experimental	High	Medium	Low
White	310	129 (42)*	181 (58)	88 (28)	58 (19)	35 (11)
Black	234	83 (35)	151 (65)	68 (29)	50 (22)	33 (14)
Spanish-speaking	149	56 (38)	93 (62)	44 (30)	37 (24)	12 (8)
Total	693	268 (39)	425 (61)	200 (29)	145 (21)	80 (11)

* Percents in parentheses

In addition to distinguishing generosity levels, an alternative and hopefully more efficient set of control variables is used here to isolate the experimental response. The primary objective is to "hold constant" the labor supply behavior which would have obtained in the absence of the experimental treatment, and to measure responses as departures from that ex ante standard. In other words, one aims at contrasting statistically matched groups who were eligible for varying levels of negative tax benefits or none at all in the case of control families. Families or persons can be said to be matched if they have combinations of characteristics which are associated with equal expected levels of labor supply in the absence of any experimental treatment.

For this purpose two variables were "constructed" and estimated by a "first-stage" procedure to provide control of wage rates and family income, both on a "normal" or ex ante treatment basis. A complete explanation and specification of these variables can be found in the full report on the experiment* and will not be repeated here. Very briefly the objective was to smooth the multiple observations on both wage rates and family income (exclusive of work-conditioned transfers) and to eliminate components correlated with experimental treatments. The resulting variables, $\hat{W}$ and $\hat{Y}$, each include a component that is a function of explicit personal or family characteristics and a component which reflects persistent deviations from the average relations and, implicitly, the unmeasured factors such as motivation, handicaps, etc., which lie behind these deviations. This device was developed to more fully utilize the potential of panel data for distinguishing "permanent" or "normal" earning capacities of persons or families.

---

*Final Report of the Graduated Work Incentives Experiment, Institute for Research on Poverty, December 1973, Part B, Chapter 1.

A standard functional form was adopted which uses these variables as measures of the more traditional determinants of labor supply. This form is used for each of the four labor-supply indicators. This function, $S_c(\hat{Y}, \hat{W}, PL)$, is specified as follows:

$$S_c(\ldots) = c_1\hat{Y}/PL(n) + c_2\hat{W} + c_3/\hat{W} + c_4\hat{Y}/\hat{W} \qquad (1)$$

where $\hat{Y}$ and $\hat{W}$ are the constructed variables denoting normal income and normal wage respectively. In the first term the "normal" income variable is normalized by dividing by the poverty level, PL(n); that is, a "welfare" ratio is used. The second and third terms allow for a nonlinear wage-rate effect and the fourth term allows for an interaction between wage rates and income. We have:

$$\frac{\partial S_c}{\partial \hat{Y}} = \frac{c_1}{PL(n)} + \frac{c_4}{\hat{W}} \qquad (2)$$

and

$$\frac{\partial S_c}{\partial \hat{W}} = c_2 - \frac{1}{\hat{W}^2}(c_3 + c_4\hat{Y}) \qquad (3)$$

so that the normal income derivative varies both with need, as measured by PL(n) and the wage rate. The normal wage rate derivative varies both with wage rate and with normal income.

This function, along with linear terms in pre-enrollment hours and weeks employed in base year, provided a substantial reduction in the

variance of estimated experimental response, and in cases where the latter approached significance showed basically the same pattern of response. Consequently, the coefficients displayed in Table 5 have been derived from regressions using S_c and the pre-experimental values as the means for defining the ceteris paribus comparisons. The treatment variables are defined as the Low, Medium, and High plan groupings explained above, formulated as dummy variables.

We find in this table that 19 out of 144 individual response coefficients are significant at a 10 percent level or better. This is slightly more than a random data set would provide. Of the joint tests in the last two columns five out of 48 are significant for the test of a null response and only two out of 48 reject the hypothesis of no effect of generosity.

In spite of the scarcity of significant results it may be useful to review the ones which do appear. It will be noted that four of the significant F_3 tests and twelve of the significant coefficients are found for earnings. Nearly all these are positive and are not explainable in terms of similar significant positive responses in hours worked. White males during the third year show significant effects on differences in employment, hours, and earnings. The results here show a relatively low participation rate for the white Low group and a relatively high unemployment experience for all three groups, which is reflected as lowered earnings and hours relative to the control group. Only in the case of the Low group does there seem to be a consistent pattern of lowered supply across the whole period. The Spanish-speaking sample displays only one significant coefficient--an excess of hours worked in the Low group in the third year. For blacks (except for the earnings results) the only significant result is the sharp contrast in participation rates between the Low and Medium groups in the third year.

Table 5

Experimental Response According to Plan "Generosity" for Four Indicators of Labor Supply by Ethnicity and Time Span

	Quarter	Base	Plan: Low	Plan: Medium	Plan: High	F_3	F_2	R^2
PARTICIPATION								
White	1-4	95.8	-1.3	1.5	4.5*	1.64	1.63	.40
	5-8	97.7	-3.8	-2.6	2.3	1.59	2.36*	.33
	9-12	94.4	-4.1	-0.7	3.9	1.59	2.31	.28
	3-10	96.9	-2.8	-1.2	2.9	1.39	2.04	.37
Black	1-4	98.3	1.8	0.7	-1.0	0.23	0.34	.25
	5-8	97.8	-3.3	2.0	2.7	1.03	1.43	.26
	9-12	99.3	-5.8	5.7	-0.8	2.05	3.08**	.26
	3-10	97.5	-2.0	2.6	1.5	0.65	0.87	.28
Spanish-speakers	1-4	99.0	3.0	0.4	-0.1	0.19	0.27	.22
	5-8	100.6	1.9	-0.2	0.3	0.09	0.12	.30
	9-12	99.3	1.9	2.2	-0.8	0.42	0.59	.22
	3-10	99.6	2.4	0.0	0.1	0.13	0.18	.29
EMPLOYMENT								
White	1-4	92.5	-1.7	0.2	2.6	0.52	0.67	.45
	5-8	92.5	-6.7*	-4.3	0.1	1.52	1.66	.47
	9-12	87.6	-9.6**	-4.7	-3.0	1.52	0.86	.37
	3-10	91.6	-6.3*	-4.2	-0.1	1.55	1.60	.50
Black	1-4	94.0	3.1	3.1	3.5	0.50	0.01	.41
	5-8	93.9	0.1	-0.9	2.1	0.22	0.30	.37
	9-12	92.9	1.1	2.2	2.7	0.15	0.04	.42
	3-10	92.6	1.7	1.6	2.7	0.26	0.06	.44
Spanish-speakers	1-4	96.6	0.1	-3.8	-2.5	0.40	0.23	.31
	5-8	95.7	3.2	-3.6	-0.6	0.35	0.47	.33
	9-12	97.0	8.1	2.3	-1.8	0.74	1.08	.42
	3-10	95.7	4.9	-2.9	-0.5	0.58	0.83	.42

Note: From regressions controlling on S_c [$\hat{Y}$, $\hat{w}$, PL(n)], and pre-experiment values for hours per week and weeks per year. The base is for head with $\hat{Y}$ = \$100, $\hat{w}$ = \$2.50, PL(n) = \$80, who worked 40 hours at pre-enrollment and for 50 weeks the previous year.

F_3 is the F-statistic for the null hypothesis that all three experimental coefficients are zero, i.e., all experimental group means are the same as the base. F_2 is the F-statistic for the null hypothesis that all three experimental coefficients are the same, i.e., no variation by generosity.

*Statistically significant at the 10 percent level.

**Statistically significant at the 5 percent level.

***Statistically significant at the 1 percent level.

Table 5 (cont.)

	Quarter	Base	Plan: Low	Plan: Medium	Plan: High	F_3	F_2	R^2
HOURS PER WEEK								
White	1-4	37.2	1.4	-1.0	-1.0	0.73	0.99	.56
	5-8	35.8	-3.9*	-1.5	-0.9	1.33	1.05	.49
	9-12	35.2	-2.5	-4.4**	-3.2**	2.48*	0.34	.47
	3-10	36.2	-2.7	-2.1	-1.3	1.29	0.37	.58
Black	1-4	36.4	3.0	2.5	1.9	1.05	0.15	.43
	5-8	35.7	1.5	0.8	1.1	0.26	0.06	.51
	9-12	36.5	1.3	0.7	1.8	0.34	0.13	.47
	3-10	35.7	2.1	1.2	1.3	0.60	0.12	.56
Spanish-speakers	1-4	37.5	-2.7	-0.8	-1.9	0.53	0.26	.38
	5-8	37.8	2.3	-0.6	-0.3	0.30	0.44	.44
	9-12	37.7	5.6*	1.9	0.9	1.19	1.15	.47
	3-10	37.4	3.1	-0.2	0.1	0.65	0.95	.54
EARNINGS PER WEEK								
White	1-4	93.0	10.4**	1.1	2.2	1.72	1.81	.63
	5-8	93.8	-6.0	-3.1	-0.5	0.63	0.61	.68
	9-12	88.4	-4.5	-13.4***	-5.8	2.51*	1.32	.61
	3-10	92.9	-3.4	-4.3	-0.6	0.67	0.58	.75
Black	1-4	95.5	8.7	11.2**	8.6*	2.24*	0.15	.52
	5-8	92.3	16.8***	10.1**	9.6**	3.54**	0.82	.60
	9-12	95.8	14.7**	8.2	9.3*	1.80	0.39	.54
	3-10	92.5	16.3***	9.8**	9.6***	5.02***	1.09	.68
Spanish-speakers	1-4	96.3	-9.6	2.4	2.5	1.28	1.87	.63
	5-8	94.7	1.3	-2.9	0.5	0.18	0.26	.59
	9-12	94.4	9.6	1.2	4.9	0.76	0.63	.64
	3-10	93.9	3.4	-1.3	3.5	0.62	0.77	.73

It is clear that this first step of distinguishing crude levels of treatment has not disclosed substantial and consistent response differentials. Rejection of the null hypothesis occurs barely more often than would be the case with random responses. The significant results are particularly scarce in the middle year and middle two-year periods where the most relevant and reliable responses might be expected. Again, the effects on labor supply, if any, are too small or subtle to be detected by these simple models.

Models with Treatment-Income Interactions

Prior to considering the next phase, two reductions in the scope of the analysis must be explained. First, the variable, weekly earnings, is discontinued as a labor supply indicator. The frequent inconsistency between the signs or significance of response in hours and earnings reinforces a previous suspicion about the reliability of experimental-control contrasts for earnings. Briefly, the observed divergence between control and experimental earners in average hourly earnings may be largely an artifact of differential rates of learning to report gross rather than net or take-home pay. This matter is considered at length in another part of the report (Part B, Chapter I), but the judgement was made here that the possibility of differential bias in the measurement of earnings was great and the consequent (positive) bias in the estimates of response could be avoided by abandoning that variable.

A second modification eliminated the sample cases in the two "low" experimental cells. As mentioned above, the families involved are overwhelmingly either above the breakeven point or, if below, have opted for more generous public assistance (welfare) benefits. Consequently, they offer little useful evidence about the effects of positive benefits or

assigned marginal tax rates which affect persons below the breakeven.

The model employed here incorporates a quite elaborate form designed to allow the relation between income and the assigned breakeven point to condition the experimental response. The asymmetrical nature of the benefit schedule and the associated marginal tax rates suggest that the response to the treatment ought to vanish at some level of earning. Presumably this vanishing point would be above the breakeven level (where net benefits stop and the marginal tax rates revert to the same schedules as the regular positive tax system faced by control units.) To capture this notion a variable, Θ, was constructed to embody the assumption that a male head who is more than 20-hours-worth of work above the breakeven point is "immune" from the effects of the experimental treatment and can be regarded as equivalent to a control observation. This variable, therefore, is identically zero for control observations and for those experimental husbands who could forego 20 or more hours of work, at wage $\hat{W}$, without falling below their breakeven income. Otherwise it is defined as

$$\theta_{it} = \frac{M_{it} + 20\hat{W}_{it} - \hat{Y}_{it}}{10\hat{W}_{it}} \qquad (4)$$

where $M_i = \frac{G_i P(n_{it})}{t_i}$ is the (dollar) breakeven level for the i^{th} household, given its assigned G_i and t_i, and the poverty level $P(n)$ appropriate for its size, n

$\hat{W}_{it}$ = the "normal" wage for the i^{th} husband at time t

$\hat{Y}_{it}$ = the "normal" income for the i^{th} family at time t

The last two components are, again, the constructed variables mentioned above. The variable, θ, will be equal to 2 for an observation with $\hat{Y}$ precisely at the breakeven level, and takes on higher positive values for cases that are

below the breakeven level.

In addition to providing an index of distance from a response "vanishing point" this variable is scaled in terms of (tens of) hours. It is proposed that such a measure has greater comparability on an interpersonal basis than would the simple dollar amount of the gap. The "normal" components are used here, however, to avoid the introduction of a possible response in income or wage rates into the explanatory variables.

Clearly θ depends on the guarantee and tax rates as well as upon income and wage rates; and it requires some care to interpret the work response associated with Θ and different levels of G and t. One cannot hold Θ constant for a given person and vary G and t freely. A mean of 5 and a standard deviation of 3 are good round numbers to describe the distribution of θ. The whites show a smaller mean and declining trend, but have more variation. The means for blacks show a rising trend and have a variance similar to the Spanish-speaking group. The latter's mean displays a weak tendency to decline.

The response function used with θ is a six-parameter, homogeneous-in-θ function as follows:

$$X = (\alpha_{11} + \alpha_{12}S_2 + \alpha_{13}S_3)\theta + (\alpha_{21} + \alpha_{22}S_2 + \alpha_{23}S_3)\theta^2 \qquad (5)$$

where $S_2 = G - .75$ and

$S_3 = t - .5$

Hence, the response is constrained to zero for $\theta = 0$ and is a quadratic function of θ for any given values of G and t. The quadratic coefficients

are, however, linear functions of G and t. The coefficients α_{11} and α_{21} give directly the coefficients for θ and θ^2 when G = .75 and t = .5. Alternative combinations of G, t can be generated using appropriate values for S_2 and S_3.

The constraint of homogeneity was tested (by introducing a constant in X) and was rejected less often than the test level would imply. The constraint of linear forms for the coefficients of θ was not tested directly, but previous attempts to get consistent nonlinear patterns across the variation in G and t have been unsuccessful and, moreover, the magnitude of apparent response is so small that six parameters already seem an excessive burden to place on the systematic variation. On the other hand, the quadratic terms have generally proven to be an important part of the function whenever there is a substantial response, so it does not appear advisable to drop that part of the function. Finally, considering that G and t are also present in θ, the overall function is not linear in these variables. (The consideration of the form of this function's derivatives will be explored below at greater length.)

Along with this form for the basic experimental response a total of 19 additional variables (plus constant) have been introduced to isolate the response for comparable husbands. Four of these variables relate to health status and a health-experimental interaction. This implies that the main experimental response applies to husbands who are in good health. The four-variable sub-function, S_c, described above, was augmented again by pre-experiment values of hours per week and weeks per year. In addition a 5-parameter, piecewise-linear (spline) function of age and education, number of adults and children in the family and the estimated average normal wage rate and family income (i.e. the parts of $\hat{Y}$ and $\hat{W}$ that are independent of

the individual's persistent deviation from the mean relation).

In combination, these variables explain a substantial amount of the variation in participation and related labor-supply variables. The R^2s range from 40 to 50 percent for participation and even higher for employment and hours. Because the focus here is on purifying or refining the control-experimental comparison there has been no extended attempt to "clean up" the control function. It contains some redundancy, and moreover, the specification of some variables, such as normal income and wages would require more work if the objective were to secure directly results on the general supply response to variation in price and income. However, the present focus is on the response to experimental treatment and on securing a substantial level of homogeneity in the comparisons among plan groups--hence, the rather profligate use of degrees of freedom to provide an inclusive basis for controlling nonexperimental factors. Again, the estimated coefficients for this control function will not be presented in this paper.

In the interest of brevity only a small part of the results can be displayed here. Besides limiting the analysis to the "high" and medium level plans, and three dependent variables, we will here limit discussion to the analysis of averages over the central two years of the experiment (Quarters 3-10). More complete results can be found in the full report. Table 6 arrays a series of relevant F-ratio tests concerning the experimental sub-function. The first column shows (for 3 ethnic groups and 3 dependent variables) the statistic for the hypothesis of no experimental response--that is, the null hypothesis that the six parameters in equation (5) and the two which measure a response for husbands in ill health are all

Table 6

F-Ratio Tests of Hypotheses on Experimental Response Regressions on Averages of Central Two Years

	All Experimental (8)	Terms Using θ^2 (3)	Tax Rate Terms (2)	Guarantee Terms (2)	Tax and Guarantee Terms (4)	R^2
White Husbands (275)						
Participation	.43	.09	.11	.44	.46	.417
Employment	1.03	1.15	.50	1.02	1.33	.492
Hours	2.40**	.86	.95	2.57*	2.58*	.637
Black Husbands (201						
Participation	.84	.45	.63	.36	.35	.442
Employment	.87	.35	.24	.45	.36	.540
Hours	1.28	1.41	2.07	1.16	1.07	.658
Spanish-speaking White Husbands (137)						
Participation	3.81***	6.53***	13.06***	9.85***	6.99***	.521
Employment	5.79***	7.20***	19.72***	15.79***	9.86***	.638
Hours	3.10***	3.28**	10.33***	8.87***	5.29***	.687

* Significant at .10
** Significant at .05
*** Significant at .01

zero. The second column corresponds to the test of non-linearity in Θ--the null hypothesis is $\alpha_{21} = \alpha_{22} = \alpha_{23} = 0$. The third tests the hypothesis that $\frac{d^2x}{dtd\theta} = 0$, i.e. $\alpha_{13} = \alpha_{23} = 0$. The fourth tests a comparable hypothesis for the guarantee, $\alpha_{12} = \alpha_{22} = 0$. The next test combines the previous two; and finally the R^2 for the over-all regression is shown.

Clearly the only strong and pervasive response was found for the Spanish-speaking part of the sample. For the whites significance was achieved for hours, and in that case the non-linear pattern of variation with θ was not significant. The Black response is persistently non-significant and, as we shall see, typically inconsistent in sign relative to expectations and the pattern of the other two groups.

Turning to the estimates, the most interpretable form is to show several points on the estimated response surface. Table 7 displays a "slice" of the response surface for each of the nine regressions described in Table 6. A central plan, 100-50, has been chosen for display, and the table shows first the level of θ at which response becomes negative, θ^*. Next are shown the absolute differential estimates for those levels of θ that span most of the range of the observations; $\theta=2$ is the response estimate for a family with $\hat{Y}$ equal to the breakeven level; $\theta=5$ corresponds to a family that is 30 hours of husband's earnings short of the breakeven level, and $\Theta=8$ corresponds to a 60-hour shortfall.

It is clear that, where significant, the estimated response is negative in the below breakeven region, and shows a curvature implying relatively greater disincentive with increasing distance from the breakeven level. The insignificant estimates are predominantly positive for

Table 7

Estimated Experimental Response Surface Evaluated
for G=1.00, t=.50 plan, Healthy Husband. Central Two Years

	θ^* = "cross point" (see note below)	θ level 2	θ level 5	θ level 8
White				
Participation	9.5	1.2	1.8	1.0
Employment	5.4	1.4	0.4	-4.4
Hours	0.2	-0.4	-2.4	-6.1
Black				
Participation	8.2	1.9	2.4	0.2
Employment	17.4	3.3	6.7	8.1
Hours	10.1	2.1	3.3	2.1
Spanish-speaking				
Participation	4.5	4.0	-1.8	-21.6
Employment	3.4	3.3	-9.4	-43.2
Hours	3.1	0.7	-3.2	-13.0

Note: Column 1 for Θ^* shows the level of Θ at which the response becomes negative

Blacks, and always quite small for whites.

The estimated relationship of response to variations among the treatment plans is shown in Table 8. Because of the complex interaction of the guarantee and tax rates in the definition of θ we show a selected set of derivatives and the corresponding elasticities of hours with respect to the basic parameters of the treatment. Because the derivatives are complex expressions it is necessary to evaluate them for specific values of treatment, family and personal variables. Table 8 shows two such evaluations for each of the ethnic groups for the regressions analyzing weekly hours. Both are evaluated for the G=1.00, t=.50 plan and for a family with a poverty line at \$80 per week. The evaluation at θ=4.4 is approximately at the mean for families with the 100-50 plan. It corresponds to a wage rate around \$2.50 and a normal family income of \$100 per week. The evaluation at θ=6 corresponds to a wage of \$2.00 and a normal income at \$80 per week on the same plan. The derivative with respect to θ is shown first. Its sign and movement show that the response is generally convex in the sense of showing accelerating disincentive with increases in θ. The second column shows the derivative which corresponds to a traditional income effect, produced here as an increment to the guarantee with no change in the tax rate. The next column corresponds to a simultaneous increase in guarantee and tax rate which holds the breakeven point at the same level (M=G/t) and is called here a "pivot" effect. It shows the consequence of increasing the benefits for all those who receive them but not increasing the level of income at which benefits begin. The fourth column shows a traditional price effect (with sign reversal) here produced as an increment to the tax rate applied to earnings. Finally,

Table 8

Selected Derivatives and Elasticities for Prototype Cases from Regressions Using Central Two Year Averages of Husband's Hours[a]

		$\frac{dx}{d\theta}$	$\frac{dx}{dG}\Big\vert_{dt=0}$	$\frac{dx}{dG}\Big\vert_{dM=0}$	$\frac{dx}{dt}\Big\vert_{dG=0}$	$\frac{dx}{dt}\Big\vert_{dB=0}$
		"Gap" Effect	Income Effect	"Pivot" Effect	Price Effect	Substitution Effect
White	$\theta = 4.4$	-.8 (-.09)[b]	0.7 (.02)	2.9 (.08)	4.5 (.06)	5.3 (.07)
	$\theta = 6$	-1.2 (-.30)	2.4 (.10)	3.9 (.16)	3.0 (.06)	5.4 (.11)
Black	$\theta = 4.4$	.2 (.02)	-1.0 (-.03)	1.5 (.04)	5.3 (.07)	3.9 (.05)
	$\theta = 6$	-.2 (-.06)	-1.1 (-.05)	-.3 (-.01)	1.6 (.03)	0.5 (.01)
Spanish	$\theta = 4.4$	-1.9 (-.22)	-3.4 (-.09)	-4.0 (-.11)	-1.1 (-.01)	-5.4 (-.07)
	$\theta = 6$	-3.0 (-.74)	-1.9 (-.04)	-7.6 (-.31)	-13.3 (-.28)	-14.3 (-.30)

[a]The prototype plan refers to a guarantee, G, of 1.00 times the poverty level and a tax rate, t, equal to .50. The poverty level is approximately $80 per week.

[b]Numbers in parentheses are elasticities assuming 38 hours as base for $\Theta = 4.4$ and 24 hours as base for the $\Theta = 6$ prototype.

the last column shows the income-compensated substitution effect derived under the constraint that benefits (B) remain constant under a change in the tax-rate.

These results conform to theoretical expectations only for the Spanish-speaking, the smallest ethnic group in the sample, and the only one which consistently displayed a significant response in terms of this model. The white sample showed a weakly significant response for hours, but the derivatives are uniformly of the "wrong" sign. It is interesting, however, that even where significant and of the expected sign, the elasticities, are relatively small. The largest of the elasticities, for the lowest income Spanish-speaking husbands, are less than 1/3--hardly a sign of extreme sensitivity.

Experimental Response Constrained for a Cobb-Douglas Utility Function

The third and final review of evidence considers a model which constrains the response to conform to a theoretical specification of the form of that response.* This analysis also does not allow for ethnic differences, either in the control or response portion of the model. The experimental response is required to be proportional to a variable $Q = \frac{G}{(1-t)w'}$, (or its square) where G and t are defined as above, and w' is an alternative estimate of the person's normal wage rate. The variable Q is derived from an optimization of a family utility function and provides an extremely economical parameterization of the treatment. The dependent variable is an average over quarters 7-10, i.e., the second of the central two experimental years. This model is estimated with a set of control variables which includes pre-experimental hours per week and weeks per year, plus age, education, health status, the reciprocal of w' and binary variables

* This analysis was carried out by Professor David Horner and is reported in detail in the full report of the experiment. The utility function specified was: $\ln U = A + a\ln Y + \Sigma b_i \ln L_i$ where U is utility and Y = market goods and the L_i represent "leisure" activities of the family mem-

for the several sites. There were 799 male family heads available for this estimate (note that these are not limited to husbands in the continuous husband-wife sample). The coefficient of Q turns out to have the expected negative sign and to be significant at the 5 percent level for a one-tailed test. The response surface evaluated the average wage rate and for the eight tax-guarantee combinations is shown in Table 9.

Table 9

Percentage Response in Hours Averaged data for Quarters 7-10
(base=34.7 hours)

Guarantee	Tax Rate .3	.5	.7
1.25	X	-5.0	X
1.00	X	-4.0	-6.6
.75	-2.1	-3.0	-5.0
.50	-1.4	-2.0	X

The average wage rate used was $2.67, and the form of response requires the differential to get larger for smaller wate rates; that is, at $1.67 per hour the responses would be 60 percent larger; and at $3.67 they would be 27 percent smaller. Extensive tests were carried out on the hypothesis embodied in the restrictions by allowing the components of Q to be entered as separate variables, alone and in combination with Q. The outcome supported the one-parameter model as against the multi-parameter alternative. Once again, and this time with a relatively simple and restrictive model, on data aggregated over a shorter period and with the ethnic groups pooled together the response seems quite small.

Concluding Remarks for Males

Overall, it would not require a determined skeptic to claim that there is no evidence of any disincentive at all for husbands. If there is one, it is small on average and is somewhat concentrated among those husbands

with extremely low wage rates relative to the total needs of their families. The income effect appears quite weak, and the substitution effect appears to become strong only when net wages become very small.

The ethnic differences are somewhat puzzling, especially the typically positive (and non-significant) response estimates for Blacks. There is evidence that the Black control group, for unknown reasons, had a relatively adverse experience over the experimental period, while the Black experimental group was more comparable to the white and Spanish-speaking experimental groups. The white sample showed significant disincentives only for hours, and even these reveal some disturbing properties. The components of the treatment effects, as shown in Table 8, do not agree with theoretical expectations, but again the effects were quite small.

Once again, and here for quite low-income groups, married men with substantial family responsibilities are found to be quite insensitive to price and income variation in the determination of their labor supply. The principal qualification is for the very poorest--low wage rates and very large families--where the combination of a drastic reduction of an already low wage rate and a decrease in the extreme pressure of need by a large percentage increment to income has apparently resulted in some reduction. Even here the estimates do not suggest total withdrawal but rather a fractional reduction, perhaps in the area of overtime or in multiple job holding.

In the end, there is a distinct absence of evidence for a sharp disincentive for male family heads; there is probably some for the poorest stratum, but the evidence is weak as to its precise magnitude or nature. The distinctive ethnic patterns which have weakened the ability to generalize present some yet-to-be explained puzzles; but, on the other hand,

the replication by time and ethnic group provides a substantial amount of discipline in the search for fundamental response patterns. It is evident that a great deal remains to be done with these data but, more important, there is a clear need for more independent replication from other experimental studies now underway to see if the tendencies noted here are confirmed.

Married Women, Husband Present

The employment experience of married women in the experiment provides a critical test of the work disincentive hypothesis, mainly because married women have relatively flexible work choices and because there are more institutional arrangements to accommodate work on a short-term or part-time basis. Their role is important because of both the expected long-run labor-supply effects under a negative income tax plan and their short-run work responses to an experiment.

In the long run, less work is expected as the equilibrium adjustment to an income maintenance plan by wives and other "secondary" workers whose attachement to the labor market is relatively weak. The alternative of homework to market work is the familiar principal reason for the greater discretion that wives have in work decisions. An abundance of empirical evidence points to their labor-supply responsiveness to changes in income and wage rates--the two relevant economic variables affected by income maintenance plans.* Moreover, the trend over time has been one of increasing labor force participation of married women.

*Although the range of income elasticities and substitution (i.e., "compensated" wage) elasticities for wives measured in these labor supply studies is rather wide, their respective averages, around -.5 and 1.0, are quite large relative to the estimates for males. (See the citations of labor supply studies of wives in Cain and Watts, 1973, and Table 9.2 in the book for some summary estimates.)

It is important to recognize, however, that the employment time and earnings of wives in poor families are already at low levels. Thus, even a sizeable percentage reduction in work might, in terms of total hours reduced per year, be less than a smaller percentage change by husbands.

The flexibility in labor force behavior of married women is especially important in studying the responses to a 3-year experiment. The short duration is probably the most serious limitation of the experiment as a test of the labor supply response to a legislated (and quasi-permanent) income maintenance plan. Work behavior of male heads of households is constrained to the extent that they have a firm attachment to a job that requires them to work a fixed number of hours per week on a full-year basis. It is not realistic to expect adult males to make a major change in their schedules, knowing that at the end of three years they would probably have to reestablish their old schedules. By contrast, the freer options that women have for working less than full time, and for working periodically--taking seasonal jobs, working for a period of one, two, of three years or less, and working in a variety of jobs at different times--permits wives to show a greater responsiveness to the experimental incentives.

The prevalence of short-term working arrangements by low income wives is supported with the sample data. Among 742 husband-wife families who reported for at least 8 of the 13 quarters, the average labor force participation rate among wives was 16 percent. However, over the entire experiment 40 percent of the wives reported being in the labor force during one or more quarters. Among the 1031 husband-wife and female-headed families which reported for at least 8 of the 13 quarters, the average labor force participation rate of the married women and female heads was 19 percent. In contrast, 44 percent of these women were in the labor force for one or more quarters during the three-year experiment.

This evidence of considerable mobility regarding labor force participation implies, we suggest, a potential responsiveness to the widely varying net wage rates the wives could earn and the varying income supplements their families received.

Application of the Basic Model: Statistical Results

In this paper only the experimental treatment effects will be shown. The estimated relationships between the labor supply variables and the other independent variables, which show no untoward surprises, are available in the full report.* The presentation of treatment effects consists mainly of tables showing regression coefficients or regression predicted values of the dependent variables (calculated for typical values of the treatment variables).

All treatment plans (Table 10). Table 10 shows the estimated effects on labor force participation and hours worked for each of the eight experimental plans, with and without an interaction term for ethnicity and treatment status, for quarters 3-10 of the experiment. The control variables listed at the bottom of the table were used in the regressions which are the sources of the coefficients displayed in this table. The reference group for all the coefficient entries is the control group of wives.

---

*There are two sets of relationships of interest. A regression model which is closest to the typical cross-section model in the literature was fit for the pre-experimental labor supply status of wives. The "conventional" collection of independent variables--wife's age and husband's earnings, wife's education, wife's wage (or "predicted wage"), the number and presence of children--had similar effects on labor supply as found with other bodies of data, if allowance was made for the restriction of the sample to poor families; that is, for the use of a sample truncated on the basis of family income, which partially truncates the sample on the basis of the dependent variable (wife's labor supply). (See Nicholson, 1973.) The second regression model used post-experimental data and pre-experimental statuses of the labor supply variables. The latter variables dominated the explantory power of the regression and many of the conventional variables were reduced to insignificance. (See the full report for more information.)

Table 10

Estimated Labor Supply Effects of Treatment for Each Experimental Plan, Labor Force Participation and Hours Worked, With and Without Interactions with Ethnicity, for Quarters 3 - 10[a]

	L-F Participation (Mean = 16.3)			Hours Worked (Mean = 3.83)		
	Tax Rate			Tax Rate		
	.3	.5	.7	.3	.5	.7
All Wives						
Guarantee 1.25	X[b]	-6.5	X	X	-1.30	X
1.00	X	-6.3	-2.4	X	-1.36	-.54
.75	-3.3	-1.7	0.9	-1.27	-.26	.62
.50	0.8	0.6	X	-.62	.82	X
(F-test significance level,[c] percent)		(53)			(77)	
White Wives[d]						
Guarantee 1.25	X	-10.4	X	X	-2.03	X
1.00	X	-10.3	-6.4	X	-2.11	-1.33
.75	-7.4	-5.9	-3.0	-2.12	-1.04	-.21
.50	-3.2	-3.8	X	-1.44	-.01	X
(F-test significance level, including ethnicity variables, percent)		(41)			(77)	

[a]Control variables used in the regression include variables representing: city-sites, age and education of wife, family size and age of children, health of husband and wife, total family non-work conditioned income, pre-enrollment family income, weeks worked by wife in base year, and labor force status of wife in pre-enrollment survey week.

[b]X = no plan assigned

[c]The null hypothesis is that all coefficients are equal to zero and the significance level reported is the probability percent of a Type I error; that is, the probabilities that we are making an error if we reject the null hypothesis.

[d]Treatment interaction-shift coefficient for blacks is +6.9 for L-F participation and 1.81 for hours. Treatment interaction-shift coefficient for Spanish-speaking: 8.3 for L-F participation and .79 for hours.

In the lower part of the table the results are shown for white wives in the experimental group, and the estimate for experimental wives in the other two ethnic groups may be obtained by adding the ethnic-treatment interaction coefficient listed in footnote (d) of the table.

The table reveals two conclusions: no significant effect among the eight treatment plans and some degree of ambiguity of the results with respect to theoretical expectations. Thus, not all cells show negative responses, even for white wives for whom the overall response was most consistently negative. As mentioned above, not all families should be expected to be sensitive to the plans. In addition to "random" shocks from a variety of sources that will divert wives and other family members from responding to the treatment, some families became eligible for and received benefits from public assistance, during which periods they received no benefits from their experimental plan. Other families had incomes above the breakeven level, and the guarantee and tax rates may not have had any relevance for them. In Tables 11-12 below we examine a model which deals with this aspect of the salience of the treatment.

The theoretical expectation of a negative income effect is supported. Examining the coefficients for different levels of the guarantee along the columns of Table 10, which hold constant the tax rate, we see a quite consistent negative relation between labor supply and the size of the guarantee. Among the plans with the same guarantee level, the effect of higher tax rates is often positive, but here the theoretically expected sign is inherently ambiguous. Holding family income and family size constant, the amount of transfer payments declines as the tax rate increases. Thus, the positive effect on labor supply of this decline in transfer pay-

ments may, and apparently does, offset the negative effect of the higher tax rate.* Although the eight treatment variables in Table 10 are not even close to being statistically significant, the signs and relative sizes of the guarantee and tax effects are found in other specifications noted below where they are significant.

Treatment effects measured by three variables, T, t, and G (results not shown). To conserve on space we will briefly discuss the statistical results of a three-variable parameterization of the treatment which amounts to an abridged version of the eight-variable specification reported above. (The results of the model with three treatment variables are quite similar to a third specification reported below.) The T, t, and G specification imposes the same marginal tax rate effect across the four assigned rates: 0 (for controls), .3, .5, and .7; imposes the same guarantee effect across the five G levels: 0 (for controls), .50, .75, 1.00, and 1.25; and allows discontinuity in the level of labor supply response as between treatment and control groups by means of the treatment-status dummy variable.

One advantage of this parameterization is its economy, since simple linear effects measure t and G. A disadvantage of the specification, T, t, and G, is that it retains the assumption that families (or wives) in the experimental group are expected to respond to positive tax and guarantee variables even when they are above the breakeven level of income. Although we are keenly interested in knowing whether families above the breakeven level will respond, their behavior may cloud the interpretation of

---

*To the extent that we adhered to a prior belief about the magnitude of the tax rate effects, we expected them to dominate income effects on the basis of the following presumptions: that income effects are biased down and substitution effects are biased up in a short duration experiment (see Metcalf, 1973); that among wives and other "secondary" workers in husband-wife families substitution effects are larger than income effects in elasticity terms--an empirical proposition based on previous research (see the empirical evidence in Chapters 1 and 9 in Cain and Watts, 1973); that among poor families substitution effects dominate income effects because we may believe wage goods are valued relatively highly compared to leisure goods.

t and G effects. (This issue is discussed below in connection with Tables 11-12.)

The main result of this specification is that a significant disincentive was measured for all wives, although this was primarily attributable to white wives. This significant experimental effect from the collection of the three treatment variables stems mainly from the guarantee (or income) effect. The experimental dummy variable is also negative, but is never statistically significant. The tax effect is usually positive but statistically insignificant. However, a "pure" tax effect, obtained after holding constant (or neutralizing) the income effect is sometimes negative and always quantitatively small. All these results and those that were found for the ethnic-treatment interactions are observable in the next model discussed, so let us take up the third and final specification.

Treatment effects measured by three variables, T, P, and t' (Tables 11-12). The main feature of this specification of the treatment variables is that treatment families with incomes above the breakeven level are assigned the same zero values of transfer payments, P, and tax rates, t', as controls; and are distinguished from the controls only by an experimental dummy, T. The model of behavior may be loosely described as one in which only families with incomes below the breakeven level (at the beginning of the experiment) are assumed to respond to treatment parameters--that is, the high tax rates, the positive transfer payments, and, implicitly, the guarantee. The experimental dummy allows relaxation of this restriction on the behavior of the above-breakeven families with respect to a general treatment status but not with respect to the treat-

Table 11

Treatment Effects for Four Periods, All Wives (No Ethnic Interactions), for Treatment Specified as T, P, t'[a] and Applying to Wives in Families with Incomes Below the Breakeven Level[b]

	A. LABOR FORCE PARTICIPATION RATES						B. HOURS					
	Coefficients of Treatment Variables[c] (Significance level in percent is in parentheses)			Estimated Effects by Tax Rate[d]			Coefficients of Treatment Variables[c] (Significance level in percent is in parentheses)			Estimated Effects by Tax Rate[d]		
Quarters	T	P[e]	t'	.30	.50	.70	T	P[e]	t'	.30	.50	.70
3-10	1.4 (65)	-2.7 (8)	-4.2 (50)	-2.6	-3.4	-4.2	.64 (50)	-.75 (14)	-1.40 (46)	-.53	-.81	-1.09
		(6)						(16)				
1-4	-1.2 (69)	-3.5 (3)	.2 (97)	-4.6	-4.6	-4.6	-.52 (54)	-.90 (5)	.34 (84)	1.32	-1.25	-1.19
		(1)						(2)				
5-8	2.4 (47)	-3.2 (7)	-4.0 (55)	-2.0	-2.8	-3.6	1.00 (32)	-.71 (18)	-1.48 (46)	-.15	-.45	-.75
		(10)						(29)				
9-12	.4 (91)	-1.0 (59)	-3.5 (63)	-1.7	-2.4	-3.1	.24 (84)	-.25 (69)	-2.04 (38)	-.62	-1.03	-1.44
		(67)						(47)				

Note: See footnotes accompanying Table 12.

Table 12

Treatment Effects for Two Periods, All Wives (Ethnic Interactions), for Treatment Specified as T, P, t'[a] and Applying to Wives in Families with Incomes Below the Breakeven[b]

QUARTERS 3-10

Coefficient of Treatment Variables[c] (Significance level in percent is in parentheses)	Treatment			Estimated Effects by Tax Rate[d]	Tax Rate		
	T	P[e]	t'		30	50	70
A. LABOR FORCE PARTICIPATION							
White	-4.7	-.3	-6.3	White	-6.9	-8.1	-9.4
	(29)	(88)	(48)			(<1)[f]	
		(9)					
Black	10.2	-6.5	-7.8	Black	1.4	-.2	-1.8
	(3)	(2)	(91)			(<1)[f]	
		(3)					
Spanish	5.0	-7.9	15.1	Spanish	2.1	4.7	5.7
	(24)	(2)	(18)			(<1)[f]	
		(5)					
B. HOURS							
White	-.46	.15	-3.03	White	-1.22	-1.83	-2.43
	(74)	(79)	(26)			(2)[f]	
		(28)					
Black	3.16	-2.23	-.67	Black	.73	.59	.46
	(8)	(<1)	(55)			(2)[f]	
		(1)					
Spanish	.34	-1.89	3.20	Spanish	-.69	.05	.69
	(75)	(5)	(20)			(1)[f]	
		(17)					

[a] T = Treatment dummy; P = weekly transfer payment based on income at pre-enrollment; t' = tax rate if pre-enrollment income was below breakeven.

[b] See full report for the equivalent statistics for "predicted" income.

[c] Significance level for each coefficient and, in line below, for all three variables.

[d] Estimates are the sum of the coefficients of T + P (evaluated at the mean of P) + t' times .3 or .5 or .7.

[e] The coefficient of P in the table is the regression coefficient of P times the mean value of P for the whole sample, which was $27 per week. (The mean values of P differ for each ethnic group and these values were not used in calculating the ethnic responses.) The value of P is obtained as the amount of transfer payment that the family would receive per week if it continued to earn the same income as at pre-enrollment.

[f] Significance level for all nine treatment-ethnicity interaction variables.

ment parameters. A sharp discontinuity in response at the breakeven level is not realistic, of course, but the model has the virtue of simplicity.*

The statistical results of the model using T, P, and t' are shown in Tables 11-12. As compared with the model using t, G, and t, the quantitative estimates of the disincentive are slightly smaller but the levels of statistical significance achieved here are actually slightly greater. In the labor force participation regressions with all wives (no ethnic interactions) shown in Table 11, the treatment effects taken as a whole are significant at the 10 percent level or lower for all periods except quarters 9-12. The "average" disincentives for all wives for the three tax plans amount to 15, 20, and 25 percent reductions in labor force participation during quarters 3-10 (they were only slightly less for quarters 1-12).** The average percent reductions in hours for the three tax plans are somewhat larger, but these are not statistically significant except for quarters 1-4.

---

*We experimented with a specification in which the tax and guarantee variables interact with the "distance" each family is from its breakeven level of income. These interactions were not very informative--statistical significance was not achieved with a simple linear interaction, and the more complicated functional forms behaved erratically over the income ranges and across ethnic groups. In another model we used the treatment family's "predicted" income to determine a P value and tax rate. Family income is predicted on the basis of pre-experiment variables. This regression prediction may represent the family's normal income better than the pre-enrollment figure, which undoubtedly includes a substantial transitory component. On the other hand, the regression prediction is of untested validity as a measure of normal income. Our results were similar to the models which use pre-experiment income. These results are discussed in the full report.

**The term "average" refers to the labor supply reduction predicted for each tax plan when the family is assigned the mean transfer payment received by all experimental families. Actually, the average payment differs according to tax plan and ethnic group, but a single mean value is used to achieve standardization.

Table 12 shows results for each ethnic group for quarters 3-10. The ethnic interactions are highly significant and reveal the sizable disincentive for white wives and contrasting near-zero or positive labor supply responses for black and Spanish-speaking wives. The reduction predicted for white wives is around 50 percent for both participation and hours.

In both tables the coefficient of the amount of transfer payments, P, is consistently the most significant treatment variable, which agrees with the finding of greater significance of the G treatment variable discussed above. However, the t' effect, unlike the t effect, is almost always negative, which implies that the "pure" or compensated tax effect has the theoretically expected negative sign, a result that was not clearly shown by the earlier formulations. For all wives (Table 11), the experimental dummy, T, is never significant. In Table 12, T is significant only for black wives, and the effect is positive and large--10 percentage points in quarters 3-10. This positive T effect for black wives is, however, offset by a P effect that is large and negative so their overall predicted response to coverage by the plans was close to zero.

Summary Discussion of Additional Results Using Different Samples and Models

A large number of statistical models were used to check on the robustness of the findings based on the three basic models reported in the previous section. In addition there are a great many interesting hypotheses to test other than the narrow question of disincentives and ethnic-disincentive interactions. There is space merely to mention some of this additional work. Work with pooled data for all 13 quarters and an examination of the effects of the presence of public assistance alternatives are discussed later. In the remainder of this section we

discuss the experimental effect on wives' earnings and as an interaction with several family characteristics.

Regression results with earnings as dependent variable

Earnings are defined as reported earnings during the survey week and may be assumed to represent the product of a wage rate times the number of hours worked in the survey week. Any treatment effect on earnings that is not already measured in the effect on hours must stem, aside from errors in the data, from an effect on wage rates. There is no strong theoretical presumption for a treatment effect on wage rates; plausible arguments may be advanced for both positive and negative effects.

The regression results with earnings as a dependent variable closely parallel those with hours. The treatment effect is negative and, measured as a percentage reduction, is about the same size as for hours or labor force participation. The statistical significance of this negative treatment effect on earnings is somewhat less than was found for labor force participation and about the same as in the regressions with hours as dependent variable.

Interaction Between Treatment Variables and Children, Health Status, and Husband's Employment Status

There are a great many personal and family characteristics that might interact with the experimental treatments in affecting labor supply. Ethnicity has already been given a good deal of attention. Additional analysis was given to the treatment interactions with the presence of preschool children, with the health status of the wife and husband, and with the employment status of the husband.*

*Welfare status is discussed separately below. Another set of treatment interactions with family income were not productive, as mentioned above in footnote * on page 47.

Among these interactions only the husband's health status showed a statistically significant effect. The health status of the wife was insignificant in its treatment interaction on the wife's labor supply, and the coefficient had an unexpected sign: unhealthy wives worked more if they were in the treatment group than in the control group. The presence of children under six also had the "wrong" sign in its treatment interaction, but the coefficient was small and statistically insignificant. Clearly, a priori expectations were that the combination of income supplements and high marginal tax rates on earnings would encourage wives with poor health and with preschool children to stay home. These expectations were not upheld. The treatment-interaction with the husband's employment status had no consistent effect on the wife's labor supply, although the theoretical prediction is that a pro-work effect of the income loss (stemming from the husband's unemployment) should be less for treatment wives because increased transfer payments will cushion the impact of the income loss.

There are, of course, many subgroups in the sample that could be defined by combinations of characteristics, and it is not feasible to test for "all possible" interactions. Much more remains to be done, however, to explore interactions that have theoretical and methodological interest.

Summary for Married Women

The basic findings of our analysis of the labor supply of married women in the experiment may be summarized as follows: statistically significant disincentives are shown for white wives with respect to labor force pariticpation but not for black and Spanish-speaking wives.

Hours worked and earnings were also less for the white and Spanish-speaking wives in the treatment group, but the differences with their respective control groups were not statistically significant. There were no significant differences in hours or earnings between black experimental and control wives, and the observed difference was usually positive in favor of the experimental group. Whenever significant negative differences occurred in these tests, the size of the reduction was large in percentage terms--amounting to a 30 or 40 percent reduction for white wives. However, the base was so low that the absolute amount of reduction was not large--3 to 6 percentage points in labor force participation rates for white wives, for example. All the negative differences are reflected in the transfer payment component of the treatment; the tax rate effect was never significant, although it was usually negative. Some interpretative comments about these conclusions are discussed in the last section of the paper.

The Family as a Whole

Having considered the response of the two primary family members in some detail, we now return to the aggregate behavior of the family.* There are several reasons for interest in aggregate family behavior. First of all, any negative income tax or similar income subsidy is likely to be defined in terms of family income and need (indexed by family size and composition). The Federal Income Tax essentially applies to the family as a composite, and certainly the experimental treatments were so defined. Second, there are additional secondary earners (or potential earners) in many families, and they have been neglected so far. Finally, there are theoretical reasons for expecting the labor supply choices of family members to be a joint or simultaneous decision rather than a collection of independent or recursively ordered choices. Hence, there is ample reason for concern about the family as a behavioral unit, and reasons for expecting its behavior to be different from a simple aggregation of the independently estimated behavior of its components.

Of course, the variable of most direct interest is family earnings because, as the main component of family income, response in terms of earnings bears directly on the cost and income-reinforcing effectiveness of the subsidy to family income. Large reductions in voluntary earning both increase the cost and offset the increment to income. Unfortunately, the likelihood of a relative bias in the earnings measurement for con-

---

*This segment draws on the work of Robinson Hollister.

trols and experimentals makes one hesitate to rely on the earnings data. Consequently, the analysis has typically been carried out for total family hours as well as total family earnings. In the case of family totals, it should be noted that if there are reductions in hours which are concentrated among the lowest paid family members there should be a relatively smaller reduction in total earnings for compositional reasons alone.

In the overview (see Table 3), it was found that the only statistically significant differentials were for white families, and even for them the overall 8 percent reduction in earnings was not significant. Black families showed a 5 percent reduction in total hours and Spanish-speaking families displayed less than 1 percent. In both of these two cases, earnings differentials were positive, but not significant. Once again, the scarcity of significant gross differences forces one to search more deeply for patterns that are obscured in the simple contrasts. Nevertheless, here we have quite equivocal evidence on earnings; the whites show a substantial 8 percent reduction and the other two groups are positive, leading to a weighted average of +1.6 percent for the whole sample of husband-wife families. Hours, on the other hand, display consistent minus signs and the sample average is nearly -9 percent.

The next comparison refers to a regression model with linear terms for both the tax rate (defined as t-.5) and guarantee (defined as G-1.00) in addition to an experimental constant.* The three terms are further multiplied by the ratio of family normal income, $\hat{Y}$, to family need,

*The actual model was further complicated by estimating part of the response as an explicit function of time. The results presented have been simplified by taking the simple average of the time spline ordinates at quarters 4 and 8 to approximate the same central 2 years discussed for the husbands and wives.

PL(n), that is, by a "welfare" ratio. The six parameter form has been evaluated at G = 1.00, t = .5, and for three levels of the welfare ratio that approximate the central values of the three income strata used in the initial design (90 percent, 112 percent, and 140 percent of the poverty line, respectively) Table 13 displays the response values in the columns labeled $\Delta H(1.00,.5)$ for hours and $\Delta E(1.00,.5)$ for earnings. The derivatives of the response function with respect to G and t are also shown, evaluated at the same three values for the welfare ratio. Overall tests of the null hypothesis of zero response were rejected at the 1 percent level in all six cases. (Note however, that the significance test refers to evidence for all three years, not the central two years.) The interactions with $\hat{Y}/PL$ were significant at 1 percent for the Spanish-speaking (both hours and earnings) and for Black hours. Black earnings and white hours were significant at the 10 percent level, and white earnings not at all.

The obvious hypothesis is that the response should diminish at higher levels of normalized income, particularly as that income exceeds the break-even level. Looking at the ΔE column we see such a tendency only for the Spanish-speaking families. The other two groups both show constancy in the earnings response, though at a level of response that is positive for Blacks and negative for whites. In the hours comparison both whites and Blacks show increasingly negative responses at the higher strata. Of course, the patterns are only shown for the 1.00/.5 plan. Other plans may show different behavior. These may be generated using the derivatives, or slopes, which are constant at a given income stratum. For example, the response in hours for a low-stratum Black family on a .50/.3 plan would be $(-1.2) + (-.5)(-5.7) + (-.2)(10.2) = -.39 = \Delta H(.50,.3)$. The tax rate slope is typically positive and the guarantee slope is usually negative at the central 1.00/.5 plan.

Table 13

Predicted Response and Derivatives for Family Hours and Earnings Evaluated at G=1.00, t=.5, and Three Levels of $\hat{Y}/PL$ (the Welfare Ratio)[a]

Ethnic Group	Welfare Ratio (Y/PL)	HOURS Predicted ΔH (1.00,.5)	HOURS $\frac{\partial H}{\partial G}$	HOURS $\frac{\partial H}{\partial t}$	EARNINGS Predicted ΔE (1.00,.5)	EARNINGS $\frac{\partial E}{\partial G}$	EARNINGS $\frac{\partial E}{\partial t}$
White							
Low	(.90)	-2.6	2.1	-7.2	-8.8	-8.1	11.2
Mean	(1.12)	-3.3	.6	-2.4	-8.7	-8.4	15.6
High	(1.40)	-4.5	-1.3	3.8	-8.6	-8.9	21.3
Black							
Low	(.90)	-1.2	-5.7	10.2	6.9	-8.6	27.1
Mean	(1.12)	-2.7	-10.2	16.2	6.9	-18.2	42.4
High	(1.40)	-4.7	-16.0	23.7	6.9	-30.4	61.8
Spanish-speaking							
Low	(.90)	-3.3	-11.8	2.3	-3.4	-2.6	3.9
Mean	(1.12)	-3.0	-5.4	11.4	-1.2	-13.7	21.1
High	(1.40)	-2.6	2.6	23.0	.1	-27.9	43.0

[a]Regression specification of treatment variables (with Z a vector of control variables):

$$H, E = F(Z) + T\left[a_1 + a_2(G-1.00) + a_3(t-.5) + a_4(\hat{Y}/PL) + a_5(G-1.00)(\hat{Y}/PL) + a_6(t-.5)(\hat{Y}/PL)\right]$$

Altogether the results are not very satisfactory either in showing a pattern conforming with expectations, or in showing any pattern that is consistent across ethnic groups.

A substantial amount of additional effort has been devoted to a search for more consistent patterns, and much of this is reported in the full report, but "success" has not yet been achieved. In particular, the predicted variability of income (based on pre-experiment characteristics) has been added as a determinant of response as well as distinguishing families "normally" above or below the breakeven level. While some of the results are less anomalous, there is still a lack of consistency. The Blacks show positive differentials in both hours and earnings (nearly 15 percent for earnings). The whites show large (16-18 percent) reductions in hours but moderate (7-8 percent) reductions in earnings. The relatively small Spanish-speaking sample is more erratic and shows moderate reductions in hours and substantial reductions in earnings.

It seems clear that further work is needed on the family analysis. The sample truncation poses particular problems here. Recalling that the response was negligible for husbands and often substantial for wives, and that two-earner families are quite likely to be close to the high end of the eligibility interval, it is not surprising that the higher stratum families may show larger reductions. Other secondary earners may produce similar effects if they show substantial negative responses. It should be remembered that this is a consequence of the sampling restriction, not a contradiction of the hypothesis that higher stratum (or above breakeven) families should respond less strongly (or

not at all).

In summary, the final results are broadly consistent with the earlier findings for the husband and wife, but there is evidence of additional response for other adult family members that is not as yet well sorted out. In addition, a more intricate stratification, based on the predicted labor supply status of secondary workers (perhaps embodying the "added worker" effect), appears to be needed to secure an adequate representation of the interaction of ex ante income level with response.

Some General Checks and Qualifications

The Effect of Coexisting Systems of Public Assistance

It is important to test whether the presence of an alternative welfare system and subsequent "loss" of both treatment and control families to the public assistance rolls distorted our findings about the effect of the experiment. Including welfare families may understate the disincentive of the income maintenance plan for the following reason. Nonworking families among both experimental and control groups are potentially eligible for and have a potential incentive to apply for and receive public assistance. However, some nonworking experimental families, particularly those in the generous plans, may well forgo public assistance, which probably carries some stigma and involves some transaction costs. By contrast, families who are in the control group with similar tastes and under similar circumstances do not have this favorable alternative to welfare. A systematic selection mechanism is at work, therefore, allocating a disproportionate share of control families that earn and work the least onto the welfare rolls. The less generous the welfare system the more labor we would expect to be supplied by the control group, and hence the larger are the disincentive effects we would expect to measure.

By eliminating welfare families, defined as families receiving welfare payments for more than two quarters, we expect to eliminate disporportionately more of the nonworking control families than nonworking treatment families. For this reason we expected to see a larger treatment disincentive. However, the evidence was mixed for this test. For wives, the slightly larger treatment disincentive we measure is inconsequential. What is more puzzling, the treatment effect was more positive among blacks wives in the nonwelfare sample.

After-the-fact rationalizations may be made, but the lack of a stronger treatment disincentive for the nonwelfare sample is both surprising and reassuring concerning the robustness of the earlier results. Perhaps even in experimental families the adults who found themselves in a situation where they could not work were likely to apply for the quasi-permanent benefits of welfare (which, let us remember, would possibly include medical and other benefits not provided by the experiment.)

Regarding the ethnic differences, it is possible that white families in the control group looked upon themselves as more "permanently poor" than nonwhite controls, that whites were more knowledgeable about obtaining welfare benefits, and that they found it easier to get on welfare than nonwhites. These differences may, of course, reflect the differences in the administration of welfare between Scranton, Pennsylvania and the three New Jersey sites. The New Jersey welfare program covering male-headed families--the only welfare program relevant to nonwhite families as far as our experimental data are concerned--was begun after the experiment was underway and sharply cut back before the experiment ended. Under these circumstances an incomplete adjustment and response to the program is understandable.

We conclude that the measurement of the treatment effect with our sample is not sensitive to the inclusion of welfare families. This is not to say that the same findings would have emerged had there been no welfare system. All states, however, have some form of welfare, and the setting in which any legislated negative income tax plan might be administered is also likely to be a setting in which some public assistance alternative is available. For this reason, we believe the results pre-

sented earlier for the full sample are more relevant than the results with the nonwelfare sample.

Results Using Cross-Sectional Data on Hours Worked, Pooled Over Twelve Quarters

The model underlying the previous regression analyses is based on a single period as a unit of time over which we have, for each regression model, a single observation per family member about his or her average labor supply. An alternative specification was tried which exploits the fact that for each unit we have as many as 13 observations, one for each quarterly interview. The potential number of observations for the full sample of 693 continuously reporting, husband-wife families is, therefore, 9,009 (the same as for the analysis of male heads), although they are not, of course, "independent" observations.

This pooled sample was used for some parts of the analysis of husband's and total family labor supply. In each case a components-of-variance model was utilized and estimated weights were used to correct for the non-independence of successive observations on the same unit.

For husbands, the pooled sample was used to estimate the experimental response in hours conditional on employment. Only for the white sample was there a significant (negative) response. Further work is needed to divide that effect into overtime, second job, or straight-time components. For the family totals, the pooled analysis was used to develop a time profile of response, some parts of which we reported above.

The pooled sample was also used to advantage for wives. The large number of observations permits us to estimate a treatment response by concentrating on the working wives. We thereby avoid fitting the regression to a set of observations in which the modal group has zero

hours worked and the remaining observations are distributed at lesser densities in an irregular pattern over the positive range. The model of labor supply is one in which the estimated hours-of-work function is conditional upon an affirmative labor force participation decision.

Our results using this model showed no significant treatment effects in this conditional regressor. This null finding may simply reflect more errors in measurement concerning hours relative to labor force participation. A second possible reason is that the wife's decision about labor supply is exercised only at the point where she chooses to enter or not to enter the labor force. Institutional constraints may interfere with her desired choice of the number of hours worked, given the decision to work. Or we may simply not be capturing the determinants of her hours worked if they are more fully under her control.

Conclusions and Interpretations

Overall the labor supply response to the negative income tax plans was quite small. Certainly, for Black families there was no negative effect. For white and Spanish-speaking families the effects are negative, sometimes significant, but not very large. They consist of a reduction in hours of white male heads, an increase in the unemployment rate of Spanish-speaking male heads, and a large relative reduction in the labor force participation rate of white wives.

Altogether the results are not surprising unless one had an expectation of drastic and immediate responses to what were, on average, relatively small increases in income. The literature, at least with hindsight, provides ample precedent for concluding that the work patterns of male family heads are relatively inelastic with respect to changes in income and wage rates; and that the work patterns of females, especially mothers, are more sensitive because their work habits and opportunities are more flexible. The experiment and its analysis to date have raised many new questions and are subject to many qualifications; however, we suggest that the burden of proof has been shifted to those who assert that to supplement the incomes of poor families through a negative income tax or similar redistribution mechanism would result in large labor supply reductions.

These findings have shown a reasonable degree of stability after several checks for biases. In addition to the alternative specifications and checks discussed in the previous sections, we can report that the initial examination of the problem of attrition has not produced evidence that the limited negative treatment effect would have been very different

if all the original families were retained. On a priori grounds we expected attrition to overstate the disincentive, since low earning families among the treatment group had the most to gain by staying with the experiment.

Assuming that these and other similar "procedural" challenges to the validity of the experimental results are satisfactorily addressed, a larger question remains: Can the results of the experiment be generalized to apply to the behavior of a broader population of low income families under a legislated negative income tax plan? One issue is whether the work experience over the three-year period represents the experience of the first three years of a legislated plan. The second is whether the three-year experience represents a long-run equilibrium response to a plan that has been in operation for some time. Obviously, neither question can be answered with certainty, and the second question is more problematic than the first.

Tests for the biases due to the short duration of the experiment remain a perplexing challenge. Qualitatively, we expect a downward bias in the absolute size of the (negative) income effect and an upward bias in absolute size of the (negative) substitution effect for the transfer payment and tax variables, respectively. Work on measuring these biases is underway, but because such hard-to-measure variables as subjective discount rates and time-horizons are involved, we are not confident that much progress can be made until more data from other income maintenance experiments are available.

There are several reasons for downplaying the short duration issue. One is simply our current doubts that male heads would greatly reduce their labor supply in response to a permanent plan. A second is the empirical fact of the transitory incidence in annual poverty for intact

families headed by a nonaged able-bodied male, particularly in the relatively high wage areas of the urban North. For these families a permanent income maintenance plan does not mean permanent annual receipts of transfer payments.

Finally, there is the more direct evidence of the work behavior of wives. Despite the relatively low levels of labor force participation by the wives in the sample, there was considerably mobility in and out of the labor force, which leads us to conclude that the short duration of the experiment was not a dominant reason for seeing a distorted response by them during the three years. The pattern of their labor force behavior and the type of labor market they interact with appear to permit adjustments by the wives in their market work allocations in response to negative income tax plans.

A second issue is that our findings are in some respects at odds with the non-experimental research findings. In the experiment the zero, even positive, response by Blacks is the most notable puzzle, and, more generally, a marked disincentive was not pervasively significant for different measures of labor supply. Among wives the disincentive that was found was by and large restricted to white wives. In research using non-experimental data, fairly large and significant income and substitution effects in the labor supply functions of wives have been the rule, although these estimates do span a disturbingly large range.

Because research with experimentally generated data is so new, our suggestions for reconciling the conflicting results with conventional research must be tentative. One characteristic of wives in low-income families helps to explain why the experimental disincentive is no larger than it is; namely, their low levels of participation relative to nonpoor wives. Families with working wives are simply less likely to have incomes low enough to be

eligible for inclusion in an income maintenance plan. Given a lower bound of zero and the already low levels, there is not a great deal of "room" for further downward adjustments.

Another source of minimal disincentives for poor family members in general is that low income, husband-wife families in such relatively high-wage areas as New Jersey and Pennsylvania have incomes that fluctuate around the "breakeven levels," rather than around very low or zero levels. Because these husband-wife families do not normally have incomes low enough to make them eligible for large transfer payments, the saliency or relevance of the income maintenance plan is reduced and so is, therefore, the plan's theoretically expected disincentive.

Another reason for a weak disincentive is the likelihood that these low-income families place such a high marginal utility on income (relative to non-market activities) that the family members will work when the opportunities arise even in the face of a low net wage. Finally, the existence of alternative welfare plans does serve to attract the lowest stratum of poor families which are least able to function in the competitive labor market. The remaining strata of families above this level almost by definition exhibit a greater capacity for work and earnings, and they are less likely to show pronounced differences--at least, the extreme of zero work is unlikely. Perhaps a long-run equilibrium response, associated with a gradual learning process and "legitimation" of the receipt of transfer payments, would be different. The "long-run" is, however, also associated with rising real wages, which serve to diminish the saliency of the tax and guarantee parameters that take effect only when family income drops below breakeven levels.

Let us recapitulate our suggestions for why the labor supply function may be so inelastic for poor families: the tendency for impermanence of

the families' eligibility status, the prevailing low levels of labor supply, the low trade-off between any work reduction (greater "leisure") and lower money income, and the alternative of public assistance for those whose labor supply change is extreme in a negative direction.

The patterns of labor supply response are not as clear as we expected. Nevertheless, they call into question many of the inferences from non-experimental research about the labor supply effects of income maintenance plans. It is important to remember, however, that no empirical study has measured labor supply responses by the working poor to negative income tax plans.* The expectation of a significant labor supply reduction is based, therefore, on inferences from one type of sample and one type of process by which incomes and wage rates were generated to a somewhat different sample and an entirely different process. We have come to the conclusion that the quantitative estimates of income and substitution effects have less stability with respect to the processes generating the income and wage variation than had been believed. Even in the studies with non-experimental data, there is reason to believe, and some evidence, that income effects on labor supply will be different depending on the source of the income. Thus, the income and wage effects stemming from negative income tax plans for working poor families may well yield much smaller responses than those stemming from the normal workings of the market for middle income groups. We conclude, with some reluctance, that such inferences require models which pay closer attention to specific sources and processes of the price and income variation in the data.

There are, of courses, many unresolved issues to be pursued. The

*Glen G. Cain and Harold W. Watts, Labor Supply and Income Maintenance, A Markham book from Rand McNally Co., Chicago, 1973.

Final Report of the New Jersey Experiment really marks the beginning of research efforts--including those from other experiments and the critical review of the larger research community. All the data from the experiments will be made available to facilitate this review and further analysis.

REFERENCES

1. Glen G. Cain and Harold W. Watts, Labor Supply and Income Maintenance, A Markham book from Rand McNally Co., Chicago, 1973.

2. Final Report of the Graduated Work Incentive Experiment in New Jersey and Pennsylvania---Table of contents listed below.

PART A. AN OVERVIEW OF THE LABOR-SUPPLY RESULTS AND THEIR SIGNIFICANCE FOR CURRENT POLICY; Albert Rees

PART B. LABOR-SUPPLY RESPONSE

Chapter I — Concepts Used in the Central Analysis and their Measurement; Harold W. Watts, et al

Chapter II — (a) Labor Supply of Married Men; Harold W. Watts
(b) The Impact of Negative Taxes on the Labor Supply of Low-Income Male Family Heads; David Horner

Chapter III — (a) The Labor Supply Response of Married Women; Glen Cain, et al
(b) Relationship of the Female Labor Supply Characteristics of the Experimental Sample to Those of Other Samples; Walter Nicholson

Chapter IV — School Enrollment and Labor-Force Participation Among Young Adults; Charles Mallar

Chapter V — (a) Labor Supply of the Family; Robinson Hollister
(b) The Effects of the Welfare "Bias" on Family Earnings Response to the Experiment; Robert Avery

Chapter VI — Wage Rate Response; Harold W. Watts and John Mamer

Chapter VII — The Effect of Negative Income Tax Payments on Job Turnover and Job Selection; Seymour Spilerman and Richard Miller

Chapter VIII — The Effects of Health on the Supply of and Returns to Labor; David Elesh and Myron J. Lefcowitz

Chapter IX — Social Psychological Characteristics and Labor-Force Response of Male Heads; Sonia Wright

Chapter X — Information Levels and Labor Response; Jon H. Knudsen, Robert A. Scott, and Arnold R. Shore

PART C. THE VALIDITY AND GENERALIZABILITY OF RESULTS

Chapter I — The Problem of Attrition; Jon K. Peck

Chapter II — The Effects of Welfare on Experimental Response; Irwin Garfinkel