Larry L. Orr,¹ University of Wisconsin

I. Introduction

The initiation of the graduated work incentives experiment being conducted by the Institute for Research on Poverty and MATHEMATICA in New Jersey and Pennsylvania² has raised the prospect (some might say the specter) of an ambitious program of experimentation in social programs in general, and income maintenance programs in particular. Indeed, the enthusiasm for this relatively novel technique in some quarters threatens, at times, to outrun the capabilities of the embrionic reservoir of experience and expertise in this largely untried methodology. In response to such pressures and in the interest of rational policy formulations in the area of income maintenance, a group of researchers at the Institute for Research on Poverty of the University of Wisconsin have, at the request of the U.S. Department of Health, Education and Welfare, devoted the last few months to the development of an overall research strategy in this area. Its scope of research has encompassed a wide range of program alternatives, program effects, and research methodologies. The discussion in this paper draws heavily upon the work of this group, although it is in no sense a comprehensive report of the group's activities; in particular, I have limited my remarks to consideration of income maintenance experimentation, although an important part of our work for H.E.W. has involved the delineation of non-experimental research strategies.

It is important to understand from the outset exactly what "experimentation" means in this context, and to distinguish experimentation from the related, but distinctly different, concept of "demonstration." An experiment attempts, through the exogenous manipulation of the environment facing various economic, social, or political decision-making units, to measure their behavioral responses to variations in a particular program or program feature. Viewed in terms of a multiple regression model, the experiment seeks to generate data for the estimation of the response (dependent) variable as a function of a number of independent variables, some of which are policy parameters which can be manipulated experimentally. To achieve this goal, the experiment must obviously include several different "treatments," (at a minimum, one "experimental" treatment and a "control" or "status quo" treatment), in order to obtain estimates of differential responses. For example, the New Jersey experiment includes nine distinct treatments (including the control group), defined by different combinations of the income guarantee and the "special tax rate" under a negative income tax, with the objective of estimating the earned income response surface over the guarantee-tax rate plane. The importance of such experimental variation is that it yields information about behavioral responses

to a variety of possible program variations, both those included in the experiment and, by interpolation or extrapolation, others not included.

By contrast, what I shall call "demonstrations" typically involve little or no experimental variation of policy parameters. A uniform treatment is applied to a specified group or geographic area, often without even any attempt to define a comparable control group. Thus, it is difficult to rigorously test hypotheses in a demonstration; at best, one gets a qualitative feel for the consequences of the single program variant applied, and some idea of the administrative feasibility of the program.

While demonstrations of this type may be useful for certain purposes, in this paper I shall confine my attention to experiments in which hypotheses relating to behavioral responses to specified policy parameters can be rigorously posed and tested.

II. Criteria for the Selection and Design of Experiments

Experimentation in income maintenance is an extremely expensive research undertaking, not only in financial terms, but in terms of research talent, which at the moment is a very scarce resource in this area. It follows that experiments should be used sparingly and be carefully designed to maximize their informational output. A prime criterion for the selection of a particular hypothesis for experimental testing, then, is whether that hypothesis can be adequately tested by non-experimental (and, therefore, generally less expensive) methods. If relevant nonexperimental data exist, the presumption is against using the experimentation to generate new data. The other side of this coin is, of course, that even if non-experimental data is not available, experimentation is feasible only if it can be reasonably expected to provide a definitive test of the hypothesis in question.

For those questions where both of these conditions are satisfied, one is faced with the problem of assigning research priorities which rank experimentable hypotheses according to some set of criteria. I would suggest that the following criteria be applied, roughly in the order presented.

1) Policy relevance. The over-riding objective of research on the field of income maintenance is to provide guidance to policy makers in the revision and modification of income maintenance programs. Therefore, an obvious and appropriate criterion in developing experimental research is the usefulness of the information to be obtained from such research as an input into the policy decision-making process. This does not mean catering to political whims or pressures. It simply means that certain behavioral responses will bear more heavily upon the desirability of any particular income maintenance plan than will other responses; ceteris paribus, these responses should receive higher experimental priority. The focus of the New Jersey experiment upon work effort response is a case in point; clearly, the response of recipients' earned income will have a major impact upon the cost of a negative income tax, as well as its political acceptability in terms of the dominant Puritan ethic. Moreover, because this response might be expected to be most severe among families with male heads in their working years, the New Jersey experiment was limited to that population.

2) <u>Replicability</u>. As a second criterion, I suggest what Hollister and Cain have called the "replicability criterion."³ This criterion would restrict experimentation to those program features which can feasibly be replicated on a national scale. Any number of programs can be devised and instituted on an experimental basis which, because of their cost or administrative complexity, could not reasonably be considered to be feasible alternatives for national policy. For example, one might hypothesize that a very intensive job training and counseling program would be an effective offset to the work incentives of cash transfers. It may well be, however, that even if such a program could be carried out experimentally, it would simply not be feasible to provide such services to all recipients of a national transfer program. A corollary of the replicability criterion is that to obtain valid estimates of the effects of a national program, it must be possible to replicate the hypothesized feature of the national program in the experimental setting.

3) Adequacy of existing theory and measurement techniques. In certain areas, the current state of the art severely circumscribes the possibilities for experimentation. In many cases, we have a vague notion that a particular policy parameter may have important behavioral effects, but we have only a general idea of the mechanism by which the effect will operate, and possibly only very crude quantitative measures of the effect itself. For example, one such area is the whole question of the effects of income maintenance on the institutional structure of the community. It seems plausible that large-scale income transfers will have important effects upon the interactions of recipients and nonrecipients within a whole range of economic, social, and political institutions. Yet we really have no coherent theory of community which would allow rigorous formulation of hypotheses for experimental testing of these effects. We don't know which policy parameters would be crucial to any given effects, and therefore should be varied experimentally, and-perhaps more importantly--we don't know which non-policy variables must be controlled for in order to assure valid inference.

By contrast, underlying the investigation of work effort response in the New Jersey experiment is a well-developed body of economic theory relating labor supply to wage rates, income, and other family characteristics. Where even the rudiments of such a theory are lacking, the wisest strategy is probably to devote our research effort to develop the basic theory before proceeding to experimentation.

Closely related to the theoretical underpinnings of experimentation is our ability to measure behavioral responses quantitatively. Our ability to pose meaningful, testable hypotheses, and to generalize experimental results is severely constrained by the sophistication with which we can measure responses. Returning to the example of community effects, we face the difficult problem of defining and measuring institutional change. Can we really define appropriate indices of political participation and tension, social adjustment or alienation, or adequacy of social services? If not, the results of experimentation are likely to be ambiguous at best.

The requirement of an adequate theoretical and measurement capability is especially important for the specification of an efficient sample design. The question of adequate sample size will be discussed later in this section; at this point, suffice it to say that efficient determination of sample size depends upon our ability to predict the "normal" variability of the response variable, given the values of relevant non-experimental variables (which must be specified by our theory or empirical data.) The sample must be sufficiently large to distinguish the impact of the experimental variables from this residual noise in the response variable.

The criteria just presented bear upon the selection of hypotheses for experimental testing. Once it has been decided that a particular hypothesis requires, and is amenable to, experimentation, the question of optimal experimental design arises. This is a complex question, and this is not the place for a detailed discussion of the statistical intricacies involved. However, it seems useful to establish certain guidelines for design which are relevant for the development of a comprehensive research strategy involving a number of separate experiments; the following include some of the more important considerations.

1) Experimental objectives. While the information obtained from any one experiment may be useful in analyzing a wide variety of questions, it seems most efficient to focus each experiment upon a single dominant response variable. This limitation is imposed by the necessity of defining a single experimental objective function which is to be maximized through the sample design. Maximization of the objective function is roughly equivalent to minimizing the error of estimate in the predicted response variable; it is, essentially, the efficiency criterion for the sample design. If an experiment attempts to focus on more than one objective, it is not at all clear what the criterion of efficiency in response estimation

should be.

This should not be construed to rule out the collection of data relating to a wide range of behavioral responses in each experiment. Indeed, one of the great virtues of experimentation is that it provides a rich source of longitudinal survey data on low-income households. Still, it seems most efficient to focus upon a single over-riding objective in setting the sample design. For example, the design model for the New Jersey experiment is based on the estimation of work effort response, even though the quarterly interviews are designed to elicit additional information on a wide variety of attitudes and behavior, ranging from family expenditures to political participation and social integration.

Selection of a single objective variable for each experiment has the additional advantage of permitting the selection of a relatively homogeneous sample population comprised of those households which seem most appropriate to the hypothesis being tested or which are most relevant to policy considerations. Homogeneity of the sample is desirable on several grounds. First, given financial constraints on sample size, it seems wisest to concentrate on that type of family for which a significant response seems most likely and/or important on policy grounds. Second, in many cases, it is not clear that a single functional form of the response function would be appropriate to the behavior of diverse family types.

Finally, concentration on a single objective allows the duration of the experiment to be tailored to the particular response variable under investigation. Most of the experiments currently underway or being contemplated entail payments over three to five years. A time horizon of this length is probably quite sufficient for the investigation of, say, work effort, which may be expected to respond to fairly shortterm changes in income and wage rates. Other behavioral responses, however. may be determined by much longer-run income concepts; the retirement decision of older workers and family planning decisions are cases in point. To obtain valid estimates of these responses, a much longer payment period is probably required,

2) Comparability among experiments. One of the greatest potential values to be secured from a coordinated, national approach to experimentation in income maintenance is the ability to ensure comparability of the data collected in the various individual experiments. As noted above, experimentation can provide a rich source of cross-sectional and longitudinal data on the poor--a group for which existing data is notably meager. To be of greatest value, however, it is important that data from each of the projects be gathered on a comparable basis, so that it can be pooled for analysis. In some cases, (e.g., in the measurement of attitudes, motivation, aspirations, and the like), this will simply mean that the same interview questions should be asked in each experiment. More importantly,

however, it means that the basic economic concepts, program features, and administrative arrangements should be held constant over all experiments, unless there are explicit reasons to the contrary. In essence, I am suggesting that a uniform set of "rules of operation" (the experimental equivalent of the statute governing a national plan) should be applied to all experiments. These rules would cover such things as the definition of the family unit and family income, filing and administrative procedures, and the timing of payments. Uniformity of operating rules would not, of course, preclude variation either within or between experiments of those policy parameters, such as tax rates and income guarantees, whose effects are to be studied experimentally. Uniformity would simply control for unwanted variation in those program features which are not of experimental interest, but which might act to confound the experimental results.

It should be emphasized that any one of the rules of operation might be selected as an experimental policy parameter. For example, one might wish to study the effect of variations in administrative arrangements (filing requirements, handling of claims, agency-beneficiary contracts, etc.), or of variations in the frequency of payments. The point is, that any variation either within or between experiments should serve some well-defined experimental purpose.

3) <u>Sample size</u>. Income maintenance experiments are an extremely expensive undertaking, relative to traditional social science research techniques. In the New Jersey experiment, the payments to households in the experimental group are averaging over \$1000 per year, and even the control group families must be compensated for their time and trouble. Obviously, there is a premium on selecting as small a sample as is consistent with reasonably accurate response estimation, in order to maximize the information obtained from limited research funds.⁴ As indicated above, <u>a priori</u> theory and empirical knowledge are extremely important in determining the minimum required sample size.

The approach which has been developed for solving the sample size problem for the 0.E.O.-Poverty Institute rural negative income tax project is essentially an analysis of variance framework. The analysis requires an <u>a priori</u> estimate of the "normal" variance of the response variable (i.e., the variance in the absence of the transfer program), given the values of other relevant characteristics of the response unit. For example, if the response variable is family earned income, one would want to control such family attributes as education, occupation, family composition, etc., in estimating the year-toyear variance in family earned income.

Suppose now we wish to ask the question of whether the transfer has any significant effect on the response variable. One way of posing this question is to ask whether the difference between the mean response of the experimental group and the mean response of the control group is significantly different from zero, at some specified

confidence level. For any given size of control and experimental groups, one can compute the "normal" variance of this difference and estimate the range of the response variable which falls within the specified confidence interval--i.e., the minimum response differential which can be detected with control and experimental samples of this size. If, for example, the standard error of family income, given family characteristics, is \$600 per year, the standard error of the difference between control and experimental samples of 300 families each would be \$69. This means that an observed difference of \$136 per year would be significant at the 95% confidence level. Alternatively, this means that if aver-\$4000 per age family income in the sample year, a total sample of 600 families would be sufficient to detect a difference in earned income of about 3.4%.⁵

The prime requirement is, again, adequate a priori theory and empirical knowledge. The more accurately we are able to estimate the response variable in the absence of the experimental treatment, i.e., the smaller the residual error variance, the smaller will be the sample size required for any desired degree of estimation precision. Second, given the best available estimates of the normal variation of the response variable, a decision must be made as to the precision with which we wish to estimate the response. This latter decision is obviously conditioned upon the importance of the response for policy formulation; the smaller the sensitivity of policy considerations (program cost, for example) to the response variable, the larger will be the minimum level of response detection which can be tolerated and, therefore, the smaller the required sample size.

III. Priorities for Experimentation

Taking as our starting point the OEO work incentive experiments in New Jersey and the rural areas of Iowa and North Carolina, a number of possibilities for further experimentation suggest themselves. In this section, I shall apply the selection and design criteria presented above to obtain a priority ranking of those hypotheses which seem most amenable to experimental research Heading the list of experimental objectives are a variety of issues in the broad areas of work effort response and changes in family structure. The experimental possibilities in these areas seem well within our current capabilities. A lower priority is assigned to experimentation focussed on the effects of income maintenance on community structure because, although there are a number of important issues in this area, experimental resolution of these issues does not seem feasible at this time.

In discussing experimental priorities, it is important to define the major program features, or policy parameters, which characterize any income maintenance plan. The characteristics which I consider most basic include:

a) the income guarantee; i.e., the payment which a family would receive if it had no other income. In general, this payment may be adjusted for family size and composition, and the schedule of guarantee adjustments will be an important factor in assessing certain behavioral responses;

b) the implicit tax rate; i.e., the rate at which the basic guaranteed payment is reduced as family income from other sources rises;

c) the definition of the family unit in terms of who may be included as dependents, who must be included as dependents, and who may qualify as a head of household;

d) the definition of family income and the accounting period over which income is measured for purposes of determining current payments; and,

e) coordinate programs which do not involve cash transfers (e.g., in-kind transfers such as job training, day-care facilities, and social services).

It is felt that the policy parameters listed here are general enough to characterize nearly any of the income maintenance programs currently receiving serious considerations. In general, there are three basic types of programs which have been widely advocated: negative income tax plans, children's allowances, and various modifications of the existing categorical welfare programs. Each embodies a particular guarantee schedule and tax rate, defines the family unit and family income in a particular way, and may be coupled with various coordinate programs. Therefore, rather than focussing upon program types per se, it seems preferable to analyze behavioral responses to these more general policy parameters. Proceeding in this manner, the hypotheses which should receive highest experimental priority are as follows.

1) Work effort response. The crucial dependence of program cost and the possibilities for the eventual eradication of poverty through income maintenance make the work effort response of recipients a question of highest research priority on grounds of policy relevance. Moreover, the crucial policy parameters of any national program which may be expected to influence work effort (guarantee schedules, tax rates, and the income accounting period) are features which are readily amenable to replication in the experimental setting. Finally, the existing theory and empirical knowledge of labor supply provide a sound basis for the design of experiments in this area.

a) The labor supply of families headed by non-aged males would seem to be adequately covered in the existing 0.E.O. experiments. Further experimentation should focus on the work effort of the two other principal types of poor families, those with female and aged heads. The rural experiment will, of course, include some female and aged heads, but these small subsamples should be augmented with further experimental observations, especially in urban areas. These experiments would be closely patterned after the New Jersey design, in terms of treatments, sample size, sample allocation, and duration of payments. Since the normal work effort of such families is likely to be lower than for male-headed households, however, the cost of these experiments may well be somewhat greater than in New Jersey.

b) The work effort response of the aged (and near-aged) raises a unique problem which may not be amenable to the kind of experiment developed in New Jersey. Since an income maintenance program of the negative income tax type would constitute an assured retirement income which would be available at any time, such a plan might have a significant effect upon the age of retirement, especially for relatively low-income workers. Unfortunately, payments over a period as short as three years are probably not a sufficient inducement to elicit a reliable measure of the retirement response. It may be necessary, therefore, to select a sample of older workers who would be guaranteed income maintenance payments over the rest of their lives, to obtain a valid estimate of the effect of a permanent national program. A preliminary analysis of the response could be made after a fairly short interval--say, three or four years-although the experiment would continue to yield useful data for a much longer period of time.

Such an experiment would obviously be relatively expensive. However, costs could be reduced by sampling heavily at earned income levels near the break-even point, so that substantial payments would be made only to those workers who actually do curtail their work effort significantly; this would be entirely consistent with the dominant policy interest, since the majority of the retired poor presumably had incomes above the poverty line before retirement. Moreover, at age 65, Social Security benefits could be offset dollar-for-dollar against the transfers. Thus, if the sample consisted of workers in the 55-60 age bracket, one might expect to make large payments in only, say six or seven years to each household.

c) A third type of experiment in the area of work effort response which should receive high priority is the investigation of the effects of alternative specifications of the accounting period over which income is measured for purposes of determining payments. While at first blush this may appear to be relatively unimportant administrative detail, on closer examination it turns out to be a crucial feature of the transfer plan. The specification of the income accounting period has important implications for horizontal equity among households receiving the same average income over long time periods; payment levels under any particular accounting scheme could wary greatly from family to family, depending on the time-form of their income streams. In addition, the speed of response of payments to changes in family income, and therefore to emergency needs, depends critically on the accounting period; if payments are based on a lagged average of past income, as is usually proposed, they will adjust more or less slowly to changed circumstances, depending upon the length of the lag.

More fundamentally, from the viewpoint of behavioral response, the length of the accounting period may affect the recipient's perception of the marginal tax rate and, thus, his work effort. If the worker bases his work effort upon the return to labor over a fairly short time period, then a short accounting period, with rapid adjustment of payments to changes in earned income may be perceived as involving a higher tax rate than a longer accounting period with slower adjustment, even when the statutory tax rate on earnings is the same. Indeed, if this is the case, there is a conflict between the goal of making payments respond rapidly to need and the goal of minimizing the work effort disincentive.

In addition, the definition of the accounting period may create important incentives for recipients to manipulate the timing of their income stream in order to maximize payments. A short accounting period, for example, may induce greater seasonality in work effort.

These behavioral hypotheses seem important enough to merit explicit experimental variation in accounting periods, although it may be possible to integrate this investigation into those experiments which focus on the more general work effort issue. The accounting period for the majority of the New Jersey sample is a moving average of the previous three months' income with payments adjusted monthly; in addition, a small subsample will have payments based upon a moving 12-month average. The rural project is currently expected to involve three subsamples with different accounting periods: a twelve month moving average similar to the second New Jersey plan, a three-month moving average, and a fourweek average. Unlike the New Jersey plans, how-ever, the latter two plans will involve a "carryever, the latter two plans will involve a forward" of any income above the family's breakeven point, to be counted as income in later periods. This was deemed necessary to take account of the extreme seasonality of income flows typical of rural--especially self-employed-households. These plans suggest the range of possibilities in defining the accounting period. A number of further variants are made possible by the alternative methods of treating seasonal work-related and business expenses, alternative "inventory rules" for handling carry-forwards, and alternative rules for applying the carryforwards to future income periods.

The sample sizes for several of the accounting periods proposed for the New Jersey and rural experiments will probably be inadequate to give anything more than a rough indication of the impact of the accounting period specification on work effort. These and other variants should therefore be tested in other experiments, especially in urban areas.

d) A fourth type of experiment would focus on the interaction of income maintenance with manpower, job training, and other work-related programs. It has long been argued that income maintenance programs which seek to preserve work incentives should be accompanied by programs which act to enhance the employability of the poor. The recent Presidential address on welfare reform, which explicitly tied job training, employment counseling, and day-care services to income maintenance, illustrates the concern of policy-makers with this issue. The underlying hypothesis for experimentation in this area are important interactions between these two types of programs; i.e., that the r combined effect would be different from the simple additive effects of income maintenance and work-related programs taken separately. One might hypothesize that the existence of income maintenance with work incentive features increases the attractiveness of the job training, while the availability of job training or day-care serves to reduce the disincentive effects which remain in the income maintenance program.

An income maintenance experiment to test these hypotheses is already in the planning stage. It is proposed that a variety of job training and counseling services and day-care arrangements be made available to families receiving negative income tax payments, with their response to be compared to a group for whom manpower programs, but not income maintenance is available. Unfortunately, since the experimental site was selected for the wide range of manpower programs available, it will be difficult to define a control group which has access to neither type of program. However, comparison of the results with the data from New Jersey, where manpower programs are less generally available, should provide a relatively reliable control. This, of course, places a great premium on preserving comparability between the two experiments in the design of the income maintenance program. In the case of day-care facilities, which will be included in the proposed experiment, it may be feasible to define a control group which does not receive these services, since they will presumably be subsidized for the experimental group.

e) A fifth experimental possibility in the area of work effort is the replication of the New Jersey model with a dispersed nation-wide sample.⁶ This would provide a check on the generality of the results obtained in New Jersey and other experiments, by drawing observations from a variety of different environments and labor markets. In particular, by including observations from areas with high unemployment rates, one might obtain information on how work effort under a national program would vary with the level of economic activity. Moreover, such an experiment would provide observations in communities which fall between the small towns of the rural experiment and the large industrial cities of New Jersey and other urban experiments currently being contemplated, in terms of population size.

This undertaking would present some dificult administrative problems in maintaining contact with a widely dispersed sample. This might be reduced by cluster-sampling in a number of carefully selected areas, and/or contracting with a private survey organization which already has a national sampling capability. In any case, given these problems, this experiment probably should receive somewhat lower priority than some of those proposed below relating to the effects of income maintenance on family structure. Nevertheless, it has the potential for important contributions to our empirical knowledge of work effort response.

2) Effects on family size and structure. Virtually all of the income maintenance programs now receiving serious consideration provide potentially significant incentives for changes in the basic family structure of the recipients. In terms of policy relevance, these effects may well rival the effects of the program on work effort in importance. I would propose, therefore, that these effects receive an experimental priority just below the investigation of work effort response. Fortunately, experimentation appears to be feasible for at least the more important potential effects on family structure; the relevant policy parameters are readily identifiable and (at least approximately) replicable in the experimental setting, the response variables are easily quantified, and there is a substantial body of a priori empirical information upon which to base the experimental design. The following behavioral responses seem to be the best candidates for experimentation in this area.

a) Fertility. To the extent that fertility is influenced by the level or uncertainty of family income, any income maintenance program will be likely to affect family size. Perhaps more importantly, any program which adjusts payments by family size creates financial incentives to bear children. In the extreme, plans which determine payments solely on the basis of family size, such as a children's allowance, or which limit payments to families with children, as would the President's proposed Family Assistance Program, would seem to create a maximum incentive for increased fertility. Even a universal negative income tax plan would create such incentives, since it is usually proposed that the income guarantee be adjusted for family size. The importance of the fertility response is highlighted by the recent Presidential address on birth control and the long-standing (but virtually unsubstantiated) criticism of AFDC on the grounds that it fosters illegitimacy.

At first blush, it would seem that existing data might be sufficient to answer this question. A number of countries have adopted children's allowance, some (such as France's) at very substantial benefit levels. The evidence from these "natural experiments" is, however, ambiguous at best. The resulting birth rate patterns are rendered virtually unintelligible by the absence of any meaningful control group. Hence, experimentation would seem to be called for.

Unfortunately, short-term experiments in this area are unlikely to produce valid inferences as to the effects of a permanent national program. On the one hand, payments over three or five years provide a much weaker incentive to increase family size than would a national program providing payments over the entire 18 years of a child's minority. On the other hand, experimental families might be induced to shorten the spacing of their children in order to qualify for payments during the course of the experiment. It would be impossible to analytically untangle these countervailing effects upon birth rates in the experimental group.

To avoid these analytical hazards, it would be necessary to guarantee payments over a much longer period. For example, payments might be made over a period of 15 to 20 years, with adjustments in payments for any children born within that time. Analysis of results could be made after four or five years, although the experiment would continue to yield useful data for many years. This would substantially reduce, if not eliminate, both of the biases of a short-term experiment just mentioned.

The primary response rate of interest would be the birth rate, since increased family size may be expected to result in close spacing of children. Since it is conceivable that an increase in completed family size might result with no change in spacing, however, one would also want to gather data on such indicators as desired and expected family size. These would allow prediction of completed family size in the first years of the experiment, before most of the families reach their ultimate size.

The payments themselves might be structured in one of several ways. The obvious approach would be to provide treatments consisting of various negative income tax plans, patterned after the New Jersey treatments, over the entire course of the experiment. Although this would provide a valid simulation of all the major features of a corresponding national plan, it would be a terribly expensive undertaking.⁷ This approach would also raise difficult administrative problems, since the income of the recipients would have to be monitored over the entire course of the experiment. Moreover, current funding of the entire experiment would be complicated by the unpredictability of income streams, and therefore payments, over such a long time horizon.

A second approach, which would substantially reduce these problems, would involve simulating only the features of a national program most relevant to the central issue at hand. One could create the "price effect" implicit in a negative income tax with family size adjustments by simply extending a flat annual payment (again, for, say 15-20 years) to each child born during the experiment. This payment plan would be, in effect, a children's allowance for additional children born to the sample families. Although children's allowances and negative income taxation are very different in many respects, the "price" each places on additional children is very similar, at least over a wide range of family income. This may be readily seen in Fig. 1, which shows total family income under a negative income tax as a function of earned income.



Figure 1. Total and Earned Income, Before and After an Additional Child.

Line ab is the schedule of total income (earned income plus transfers) up to the break-even in-under the initial guarantee before the schedule after the birth of an additional child; the vertical distance between the two up to Y is equal to the guarantee adjustment for the child. Thus, the marginal increment to payments resulting from an increase in family size is a constant annual amount, regardless of the level of family income, so long as earned income remains below the initial break-even level 1b. Over this income range, then, a negative income tax is indistinguishable, in terms of this price effect, from a children's allowance equal to the marginal guarantee.

Unfortunately, adoption of this payment scheme would preclude experimental analysis of the "income effect" of income maintenance on fertility, since payments would be generally lower than under a full-fledged negative income tax. I would argue, however, that this question could be adequately analyzed from existing data, and is therefore not worth the added experimental cost of a negative income tax.

The cost of payments of this type could be markedly less than under a long-term negative income tax. Suppose, for example, that a sample of families with two children were selected and guaranteed allowances of \$400 per year for each additional child.⁸ The present discounted cost of a 15-year allowance, then, would be about \$3900, discounting at 6%. The average expected completed family size of families at these income levels is probably about 3.8 children;⁹ let us therefore assume that on the average, each family has two children, spaced, say, one and three years after the beginning of the experiment. The present discounted cost per family would then be approximately \$6,900.¹⁰ This cost could be further reduced if the allowances were successively reduced for the fourth and subsequent children, as they would under a national plan.

A rough estimate of the required sample size for detection of a fairly small change in birth rates or desired family size at a high level of confidence can be made on the basis of existing data.¹¹ It appears that a sample of 300 experimental families would be sufficient to detect a difference of about 10% in the birth rate over a three-year period or 4% in desired family size, as compared to the mean of the control group, at a 96% confidence level. If this level of precision is acceptable, then the children's allowance type payments described here could be financed at approximately the same total transfer cost as the current three-year New Jersey negative income tax experiment.

b) A second aspect of family structure which merits experimentation is the question of the effect of income maintenance on marital stability, in terms of divorce, desertion, and separation. Several policy parameters may be of significance here. First, to the extent that marital instability stems from economic stresses within the family, the sheer effect of additional income, i.e., the income guarantee, may serve to reduce instability. It has also been widely suggested that the schedules of adjustment of the guarantee for family size and structure which have been proposed under most negative income tax plans will create incentives for family breakup.¹² Typically, a spouse receives a smaller marginal guarantee than the head of household. Thus, the family's total guarantee can be increased if the couple separates and forms two households, with two head of household guarantees. This incentive is analogous to the incentives for family breakup embodied in the AFDC program. Under AFDC, a family is eligible for payments only if the father is not present, and it is often alleged that this promotes desertion and separation. Of course, for that group of women currently on AFDC (or those who would become eligible at some time in the future), a change to a system which provides income maintenance for intact families would reduce the incentives for separation.

Perhaps more importantly, because of the discontinuity of tax rates at the break-even level of income, a family may be able to increase its total payments by forming two households, one with income above its break-even point and one with income below.¹³ This kind of incentive would also be present in any plan which involved a non-linear tax rate, even if both of the new householdsremained below their break-even points. This effect is identical in principle to the incentive to file joint returns under the positive income tax, with its non-linear progressive rate structure.

We have only begun to consider the problems of sample design for such an experiment. Since nearly one-third of all divorces occur in the first four years of marriage, ¹⁴ it seems reasonable to select a sample of newlywed couples. It is also probable that economic considerations play a larger role in marital instability early in marriage than in later years.

While we have not undertaken a detailed analysis of the required sample size for such an

experiment, a rough idea of the requisite sample size can be obtained if we assume that divorce is a stochastic binominial process with each family in the sample having an equal probability of divorce within the experimental period. If, for example, the incidence of divorce among couples of the type selected would be 20% in the first four years of marriage in the absence of income maintenance, a sample of 900 couples would be sufficient to detect a change of 2.7 percentage points in the divorce rate at the 96% confidence level.¹⁵ In the relative terms, this is a rather wide confidence interval; it allows significant detection of changes no smaller than 13% of the initial divorce rate, with a fairly large sample. Even a sample of 2000 couples would allow detection of a relative change no smaller than about 9% of the normal divorce rate. Of course, estimation accuracy would be improved if one were able to predict the probability of divorce for individual couples more accurately than by simply applying the mean rate to all couples. Even so. it appears that a large sample will be required if we hope to detect even moderately small changes, It should be noted, though, that payments to couples with no children (or only one or two) under any given plan would be much less expensive than the average payment levels in New Jersey, where the families are fairly large. Thus, cost considerations are less restrictive with respect to sample size.

c) Split-off of dependents. Very similar to the incentives for marital instability are the incentives that may be provided for dependents to leave the family and set up new households. Again, the total family guarantee can be increased and payments may rise if the split alters the marginal tax rate facing either the original family or the new unit. The latter effect may be especially important if a dependent of a relatively wellto-do family can set up a new household with an income below the break-even point; this would be the case, for example, for teenage youths leaving families with incomes above break-even point. A variety of other dependents now living with families might also be encouraged to set up households of their own if afforded an income guarantee: grandparents, in-laws, unmarried relatives, The fact that these potential income mainetc. tenance recipients currently reside in families throughout the income distribution means that the aggregate effect, in terms of program cost, could potentially be very significant. Of course, to the extent that these dependents continue to receive support from their families after leaving, this would count as income and reduce their payments.

While it would seem to be very desirable to attempt to estimate experimentally the impact of these incentives, the design of an experiment which would constitute a valid replication of a national program poses serious difficulties. The logical experimental approach would seem to be to select a sample of families to receive income maintenance payments, and to allow any individual who leaves a sample family, and any dependents he may acquire, to qualify for payments as a separate household. This approach, however, would overstate the incentives for splitting off embodied in the corresponding national plan, especially among young adults, because it creates a "dowry effect." In the experimental setting, a young man or woman would become a differentially attractive marriage partner because he or she would be eligible for income maintenance payments while his or her compatriots would not be. Under a national plan all single individuals would be equally eligible for income maintenance, so that all would compete for marriage partners on an equal footing. This asymmetry between experiment and national program then, creates an arti-ficial incentive for others to "marry into" the sample which would tend to bias upward our estimates of the impact of the plan on marriage of dependents.

An alternative way of defining guarantees for individuals who leave the family to marry. which introduces a downward bias, might be incorporated into the experimental design in an attempt to bracket the true response under a national plan. This would involve allowing only the individual from the original sample family, but not his dependents, to receive payments; thus, again because of the asymmetry between experiment and national program, there would be an artificial scarcity of marriage partners eligible for payments, introducing a downward bias in rates of split-off. The dowry effect would still be present (in a weaker form), but if it is felt that this "scarcity" effect would outweigh the dowry effect, this subsample would provide a lower bound for the estimate of the true effect. The subsample where individuals are allowed to bring in new dependents would provide an upper bound.

For individuals or subfamilies leaving the family to live on their own, there is no problem with the acquisition of dependents. Thus, the experimental results in these cases would be relatively reliable.

We have not yet begun to explore the questions of sample size and design and program cost involved in such an experiment. Sufficient information for such an analysis is readily available, however, in existing cross-sectional data on the structure of families and subfamilies, and rates and ages of dependents leaving the family at various income levels.

3) <u>Community effects</u>. The experimental possibilities discussed so far relate to essentially individual responses of the income maintenance recipient. These can be estimated from the behavior of families in dispersed experimental samples with little or no interaction among recipients. One might also, however, expect an income maintenance program to have a variety of effects upon social, political, and economic institutions within the community as the result of interactions among recipients and between recipients and non-recipients, under a national program. To name only a few possibilities, there might be changes in the political power balance, governmental tax and expenditure patterns, attitudes toward the poor and

aspirations of the poor, and location of economic activity and the economic opportunity structure facing the poor.

To attempt to induce these effects experimentally, one would have to adopt a strategy of "saturation sampling," giving income maintenance transfers to all households in a given area who would be eligible under a national plan. Moreover, the experimental area would have to be sufficiently large to encompass the geographical extent of the particular institution or activity being studied.

Such an experiment would obviously be very expensive. There are also several reasons to question the validity of the results that would be obtained. For one thing, as mentioned near the outset, we have very little in the way of a theory of community or of institutional change upon which to base the experimental design. Moreover, it is difficult to define measures of institutional change which would serve as response variables, or even to select the policy parameters which should be varied experimentally. More fundamentally, perhaps, an experiment in a single area, measuring the program's impact on one set of institutions, would, in essence, yield a sample of only a single observation. То constitute anything more than a demonstration (albeit, perhaps, a useful demonstration), experiments would have to be performed in a number of communities. Unfortunately, in the absence of an adequate theoretical foundation, it is impossible to say just how many community observations would be required for valid inference.

Given the design problems involved in saturation experiments and the range of other important questions more amenable to experimentation, it seems best to defer experimentation in this area, and to concentrate our research efforts instead upon developing the theoretical prerequisites to experimentation focussed on community effects.

IV. A Brief Preview of Coming Attractions

To date, two experiments in income maintenance have already been funded by the Office of Economic Opportunity--the New Jersey experiment and the rural experiment about to begin transfers in Iowa and North Carolina. Since these have been described here and elsewhere in some detail, I will simply note that in terms of the priorities proposed here, they focus primarily upon the work effort response of male headed families under a negative income tax. The rural experiment will also include some families with female and aged heads, but these samples will be small and the results not necessarily applicable to similar families in an urban setting. Finally, several variants of the accounting period will be tested in both experiments but, again, the samples are small and the alternative plans are not exhaustive. These experiments may generate some useful data with respect to other questions on my priority list, but they are not designed for this purpose, so such results will probably be only suggestive at best.

To expand our knowledge in those areas of high priority not covered by these projects, the Department of Health, Education and Welfare is now considering the initiation of a small number of carefully selected and controlled income maintenance experiments, and has already given planning grants for the design phase of two experiments. The focus of these two experiments will be upon issues which rank high on the list of priorities presented here.¹⁶

The first, located in Seattle, will consider the interaction of income maintenance with various manpower and work-related programs, along the lines suggested in section III.1.d. of this paper. The cash transfer plans will be closely patterned after the New Jersey treatments to ensure comparability of the data obtained. The sample will include a substantial proportion of female heads of household, however, so that we can begin to obtain estimates of the work effort response of females. The manpower programs to be associated with the transfer payments, for at least a subsample of the recipients, will include job training, employment counseling and referral services, and day-care facilities for female heads and wives with pre-school children. To avoid the costs and administrative problems of setting up new programs and facilities the recipients will simply be guaranteed a slot in existing programs and/or a subsidy to cover the costs of using the facilities.

The second experiment for which a planning grant has been made is to be carried out in Gary, Indiana. The experimental focus there will be on the work effort response under several variations in the income accounting period. Again, the sample will include a large proportion of female heads of household. This experiment will also conform closely to the form of the New Jersey experiment. It is hoped that the transfer phase of both the Seattle and Gary experiments can begin by the end of 1970.

H.E.W. has received requests for support of income maintenance experiments in a number of cities, and will undoubtedly solicit additional proposals for experimentation in several of the areas proposed here. The outlook for a fairly ambitious experimental attack on the large remaining areas of ignorance of the effects of income maintenance is quite promising.

Footnotes

¹The research underlying this paper was supported by the Institute for Research on Poverty under a contract with the Social and Rehabilitation Services Division of the U.S. Department of Health, Education and Welfare. The author has benefited enormously from continuing discussions with a large number of members of the Poverty Institute; in particular, this paper draws heavily upon help from Robinson Hollister and Harold Watts. Needless to say, the author alone is responsible for the analysis and opinions advanced in the paper. ²For a detailed description of this project, see Harold W. Watts, "Graduated Work Incentives: An Experiment in Negative Taxation," <u>American Eco-</u><u>nomic Review Proceedings</u>, May, 1969, pp. 463-472.

³Hollister, Robinson, and Glen Cain, "The Methodology of Evaluating Social Action Programs," <u>Discussion Paper</u>, Institute for Research on Poverty, University of Wisconsin, Madison, April, 1969.

⁴An alternative method of reducing costs is, of course, to concentrate on less generous transfer plans. Given our degree of ignorance about the magnitude and functional form of the response, however, it seems preferable to stick plans which seem generous enough to elicit a significant response.

⁵These calculations were made by D. Lee Bawden and Charles Metcalf of the Institute for Research on Poverty, University of Wisconsin, in preliminary design work for the O.E.O. rural negative income tax experiment.

⁶Such an experiment was originally advocated by Guy Orcutt, of the University of Wisconsin Department of Economics, now at the Urban Institute.

⁷If payments average \$1000 per year, as in New Jersey, e.g., the present discounted cost <u>per</u><u>family</u> for a 15-year program, discounting at 6%, would be about \$9,700.

⁸Of course, experimentally, one would want to vary the payment among families, in order to estimate a continuous response function, but \$400 ought to be a reasonable average guarantee. A negative income tax which guarantees the poverty line income would carry a marginal guarantee of about \$400 for a third child.

⁹N.B. Ryder and C.F. Westoff, "Relationship Among Intended, Expected, Desired, and Ideal Family Size: United States, 1965," <u>Population Research</u>, March, 1969.

¹⁰In addition, it might be necessary to pay a flat annual allowance on the order of \$150 to all families on the experiment to secure their cooperation in keeping in touch with the research organization during the early years when most families are receiving no payments. This payment could be terminated after four or five years when the analysis of results is undertaken, so that it would add only about \$600 to the discounted cost. Such a payment would probably also be necessary under a negative income tax (in fact, this is being done in New J rsey), so it would not materially affect the comparison of costs under the two modes of payment.

¹¹The calculations which follow were made by Glen Cain, Department of Economics and Institute for Research on Poverty, University of Wisconsin, on the basis of survey data reported in C. F. Westoff, R. G. Potter, P. C. Sagi, and E. G. Mishler, <u>Family Growth in Metropolitan America</u>, Princeton University Press, 1961.

¹²For a typical discussion, see "A Model Negative Income Tax Statute," <u>Yale Law Journal</u>, Vol. 78: 269, 1968, pp. 276-278. ¹³For example, a couple with a \$3000 income, a \$2000 guarantee, and a 50% tax rate would receive a transfer of \$500 if they live together. Suppose now they split up, with the husband retaining the \$3000 income and the wife none. If each is entitled to a \$1000 guarantee (to maintain the same total guarantee in this example), the husband would now receive no transfer, but the wife would receive \$1000, increasing their total payment by \$500.

¹⁴Vital Statistics of the United States: 1964, Table 2-5, p. 2-8, Vol. III--Marriage and Divorce, U.S. Department of Health, Education, and Welfare, Public Health Service, Govt. Printing Office, Washington, 1968. ¹⁵This calculation was made in the following manner. The same variance is given by the formula: 2 p(1-p).

$$\sigma^2 = \frac{p(1-p)}{N},$$

where p is the probability of divorce for any one couple and N is sample size. For N = 900 and p = 2, σ = .0134, and the 96% confidence interval is 2σ = .0268.

¹⁶This is not coincidental, since the objectives of these experiments were defined in consultation with members of the Poverty Institute where this list of priorities was developed.