

World Bank Reprint Series: Number 121

Dov Chernichovsky

The Economic Theory of the Household and Impact Measurement of Nutrition and Related Health Programs

Reprinted with permission from *Evaluating the Impact of Nutrition and Health Programs*, Robert E. Klein et al., eds. (New York: Plenum Publishing Corporation, 1979), pp. 227-67

From: EVALUATING THE IMPACT OF NUTRITION AND HEALTH PROGRAMS.
Edited by Robert F. Kohn et al.
(Plenum Publishing Corporation, 1977)

THE ECONOMIC THEORY OF THE HOUSEHOLD AND IMPACT
MEASUREMENT OF NUTRITION AND RELATED HEALTH PROGRAMS

Dov Chernichovsky*
The World Bank
Washington, D. C.

INTRODUCTION

As a theory of choice, household economics offers a conceptual framework in which to investigate the family's responses to changes in its environment. This framework can be useful for policy-makers and planners in formulating hypotheses about the effects of intervention programs. Econometrics, the complementary statistical extension of economic theory, furnishes a versatile statistical framework for testing these hypotheses and quantifying the effects of such programs, as well as increasing our basic knowledge about the interactions between the program and their social environments.

This paper conveys, in broad terms, an economist's approach to the evaluation of nutrition and related health programs. It emphasizes the close link between economics as a behavioral science and the measurement of the impact of intervention programs. In the first section, the basic working assumptions and framework of household economics are introduced and related to the concept of an intervention program. This relationship serves to highlight the economist's conceptual point of departure in analyzing nutrition and health-related interventions and measuring their effects. In the second section, the household's behavioral objectives that are used to specify and measure the outcomes of intervention programs are discussed. These objectives provide the conceptual framework for considering various properties of outcome variables and measurement problems in section three. Section four discusses the econometric approach to socio-economic studies in nutrition, and section five discusses the basic differences between an experi-

*Mailing address: Medical, Economics, Health & Welfare Administration,
Beersheva 84 120 P.O.B. 653, Israel

mental approach and an econometric approach in measuring program impact.

INTERVENTION PROGRAMS AND HOUSEHOLD ECONOMICS

Health-related environmental programs, like malaria eradication, and largely mandatory programs, like smallpox vaccination, apparently have been relatively successful in meeting their objectives (1), while many personal health and nutrition programs appear to be less successful. A common feature of the environmental programs and the mandatory programs is that their implementation does not require individuals or households to choose how to respond. The environmental programs do not deal directly with individuals or households, and mass vaccinations may leave little or no room for individual or household choice. Programs in health and nutrition, on the other hand, often require active decision-making and behavior change at the family and individual level. Measurement of success or failure of such programs becomes complicated due to several factors including lack of data, relevant statistical tools, or most importantly, the response of the target population to the program.

Malnutrition results largely from a combination of individual and household consumption behavior and hygienic decisions and practices. This fact limits the feasibility, economic and otherwise, of mandatory and effectively controlled nutrition programs; that is, these programs are most likely to leave to individuals and households the choices of whether to participate in a program and how to use resources that become available through it. The fact that families and individuals are required to make choices makes an economic theory of the household a useful evaluation tool.

The household, whether a nuclear or an extended family, is the basic socio-economic unit that makes most decisions about investment in human beings and about consumption. 1/ The significance of a household and an individual to the community is not limited to their role as components of a sum. An individual's education, and particularly his health, often affect the well-being of others in the community; communicable diseases exemplify the interdependence between an individual and his community. This interdependence and certain cultural norms concerning the distribution of well-being among households in the community provide much of the basic rationale for health-related intervention programs.

Economists usually view the household as a harmonious microcosm that makes deliberate and rational decisions. This is a

1/ It is important to realize, however, that in some traditional cultures, tribal or village governing entities might make important decisions about investment in human capital.

basic working assumption employed to identify the systematic part of human behavior by using conceptual parameters and measurable variables. This assumption is eventually "modified" in econometric analyses by taking into account unsystematic variations of behavior. The economic analysis of household behavior can be summarized as follows. 2/ Households and individuals engage in activities to produce "ends", or consumption commodities, that have utility. 3/ These ends compete for the household's scarce human and nonhuman resources because producing more of one commodity implies producing less of others. A change in the household situational environment can change (a) the household's income or wealth, which determines how much a household can produce, (b) the commodities' relative costs or prices, which determine the relative attractiveness of different commodities, and (c) the household's tastes and preference structure. Modifications in household behavior are derived from changes, which an intervention program can promote, in one or more of these. An economic conceptualization of the household's decision-making process is sketched in a simplified manner in Figure 1.

From the viewpoint of household economics, an intervention program has at least one inherent problem: such a program "disputes" the household's ability and even willingness to realize the social consequences of its choice and to meet some specified social objectives. This is equivalent in many instances to questioning the rationality and social adequacy of the household's decision-making process and objectives. There are two critical implications of this problem. First, an effective demand for, or utilization of, program services by the target population cannot be guaranteed; second, even when adequate demand exists, it may stem from private objectives that are not congruent with program or social objectives.

Therefore, a basic requirement for evaluating program impact is to identify the "nonprogram" parameters that determine the household's demand for program services. This identification should help to indicate how much a household will use the program, when it will do so, and what use it may make of the program's resources.

2/ The approach presented here, in broad terms, is based on traditional demand theory and some extensions of this theory by Becker (2) and Lancaster (3). Differences in approach, which bears on the conceptual framework, do (and should) not affect more practical measurement issues.

3/ Those commodities can be abstract as are, for example, "good health" and "services from children".

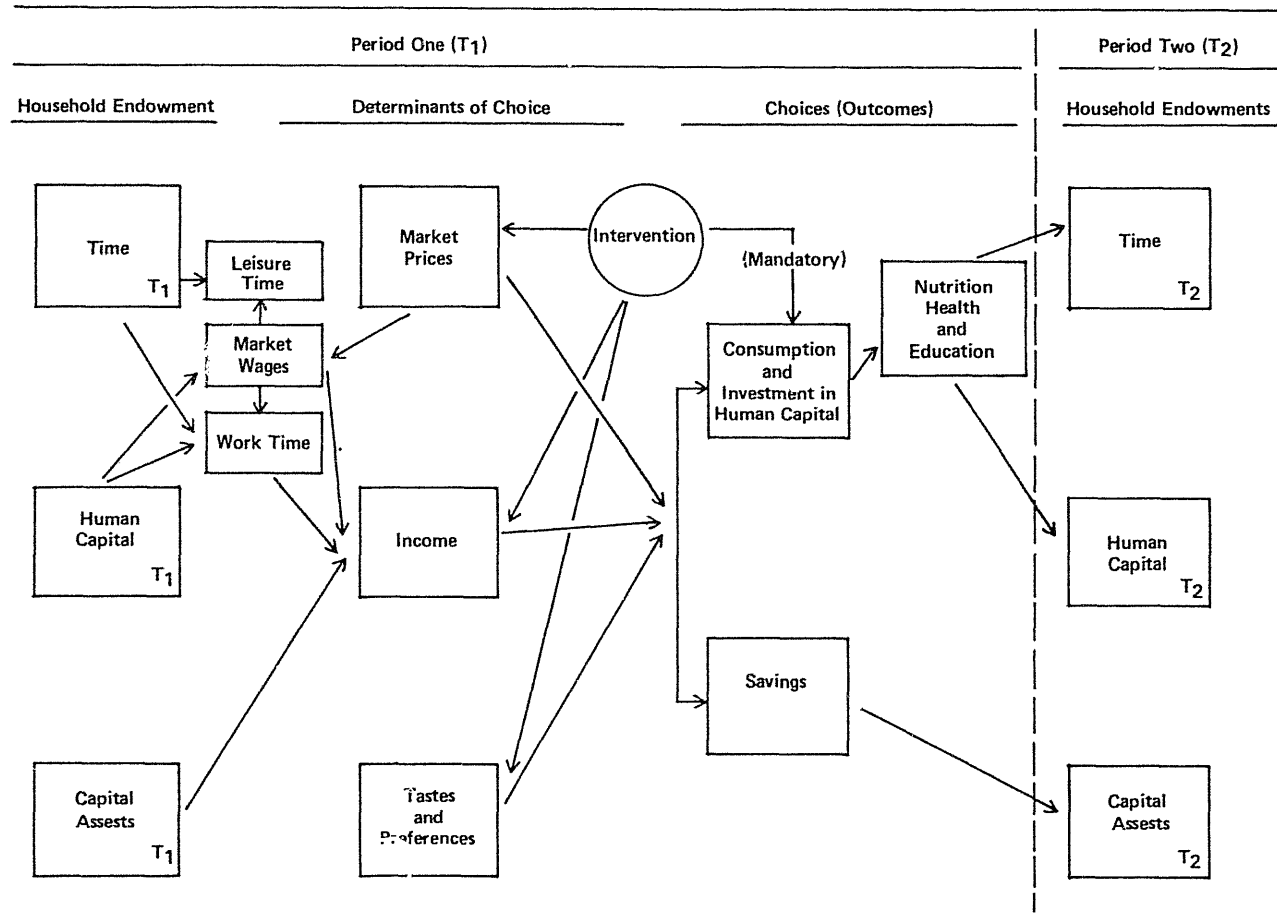


FIG. 1. An economic conceptualization of the household's decision-making process.

Program utilization is a function of (a) the degree the program serves the household's objectives; (b) the degree it draws on household resources; and (c) the relative attractiveness of subjectively conceived substitutes for program services that meet the household's objectives. These points can be illustrated by common examples. Preventive medical care programs, particularly nutrition programs, are often hard to implement because the target populations may not recognize their usefulness. In many situations, health services may compete with traditional practices that households perceive as substitutes. Program services may not be used, even when people recognize their usefulness, because of the relatively lower costs, or higher benefits, of presumed or proven substitutes, and because of the burden the program may impose on a household's immediate welfare, as when people cannot afford the time and transportation costs to go to a "free" clinic.

Program utilization need not be dependent, however, on program objectives. A target household has the capacity to reallocate program resources to reach its own rather than the program's objectives. For example, a mother may feed a particular child less than she feeds other family members because that child participates in a school feeding program. Thus, the mother may attenuate and offset the program's specific objectives and may "spread" its impact. Further, through the effect on one aspect of household life, say child mortality, the program may have an impact on other aspects, say fertility behavior. Although these other aspects may not be among the original objectives of a particular program, nonetheless, one should consider them as possible benefits or costs of the intervention.

By relating program services to household resources and objectives, the economic analysis of the household can help to identify (a) the potential uses of program services; (b) program substitutes; (c) the relative attractiveness of the program; and (d) the extent the program draws on household resources. One can thereby hypothesize who program users will be, how they may utilize the program, and subsequently, what the program's impact may be. The formulation of testable hypotheses about program impact aids in identifying variables and relationships that may measure program impact.

AN ECONOMIC FORMULATION OF HEALTH RELATED OUTCOMES

Program "impact" or "outcome" variables must be identified and discussed in conjunction with households' objectives or ends. This section deals with the household's behavioral objectives, around which we can model family responses to health intervention programs,

and on the basis of which we can identify and discuss appropriate outcome measures.

Economists have long sought to establish adequate measures of welfare, which is taken as the ultimate goal of man's economic activity. In the absence of better measures, monetary income has been used as a measure of welfare. Income, which flows from labor and accumulated stocks of assets or capital, serves as a proxy for welfare because higher levels of income mean higher levels of production and, thus, more commodities that possess utility. Measured incomes, however, ignore many utilitarian non-market household activities such as leisure. Furthermore, higher levels of income do not necessarily correlate with greater lifetime well-being when all dimensions of human welfare are considered; the health and nutrition problems of affluent societies are evidence of this last point (4).

Health status is itself a key element as well as a good proxy measure of other aspects of human welfare; it can be considered both a consumption item and an investment item (5, 6). Good health is "consumption" because it is an end in itself, accounting for a considerable part of that human welfare not measured by income. Health is also an "investment" because it is a key component of human capital that determines the level and duration of one's market and nonmarket activities. As such, it can be linked to some measurable components of earning and nonmonetary income.

This discussion focuses on the investment aspect of health because this sets the lower limit of potential benefits from health programs. 4/ The discussion is structured around the concept of expected lifetime earnings. In a simplified way, one can define for an individual of age A the present value of his lifetime expected earnings for N years henceforth by

$$E = \sum_1^N P_1^S \cdot p_1^H \cdot W_1 (1+r)^{-(i)} ,$$

where $\{P_1^S \cdot p_1^H \cdot W_1 (1+r)^{-(i)}\}$ states the present value of the earning an individual "expects" at period i . It is given by his (conditional) probability to survive to that period, P_1^S ; by the probability of his being physically able to work that year, p_1^H ; and by his average anticipated productivity -- or returns, psychic

4/ A rate of return on a health program as an investment will always understate the actual return by excluding the unmeasurable (consumption) utility derived from good health.

and other, from his time -- during that year, W_i . Given a particular time discount rate r , each time-specific expected earnings component is discounted by $(1 + r)^{-i}$, which is the present value of a unit of earning at a particular future period. 5/ Each term P_i^S , P_i^H , or W_i can be regarded as a health outcome; together they encompass the "investment" dimensions in health. 6/ P_i^S is based on age-specific mortality rates, P_i^H is based on age-specific morbidity rates measurable by lost working days, and W_i can be measured by one's average daily wage rate.

The productivity measure or wage rate, W_i , may warrant more attention because, unlike the other terms, it is not usually assumed a direct outcome of nutrition and health. In a given economic setting, defined by the available technology, land, and capital, an individual's wage rate can be taken as a function of his innate physical and mental abilities, and of his physical and mental capacities acquired through education and work experience (7). Since educational achievement and work experience are outcomes of a process that depends partially on health and nutrition, man's productivity, or W_i , can be viewed as a function of his health and nutrition, at least in situations of severe malnutrition.

Once expected earnings, as specified, are regarded as an individual's or a family's objective, the economic analysis can derive certain predictions about the household's behavior. The common approach is a maximization procedure by which the household is assumed to enhance P_i^S , P_i^H , W_i , and other utilitarian ends, subject to various constraints. This procedure sets the trade-offs among different ends, defines behavioral optima for each, and thereby provides the analytical framework to deal systematically with the relevant aspects of household behavior and to generate a set of refutable hypotheses. 7/

5/ W_i can be regarded as a term net of investment in health and education that affects the levels of all three terms.

6/ For simplicity, we ignore the interdependence among the three terms.

7/ Anthropological insights into, as well as prior evidence about, the household's view of the underlying investment process are essential for adequate modeling. That is, for example, particular household members may get substantially better diets and "health investment" than others because their (lifetime) earnings are important from the household's viewpoint. Some discrimination in feeding among household members to protect the actual or potential breadwinner is apparent in subsistence settings.

IMPACT MEASURES: BASIC PROBLEMS AND SOLUTIONS

The specification of the ultimate outcome variable -- expected lifetime earnings -- and its components is also important for discussing some key statistical issues that relate to impact measurement and to some basic properties of impact measures.

Time lags are critical in measuring program impact and present a key statistical problem inherent in intervention programs. The longer the time between the intervention and the measurement of its outcome, the harder it is to link the program to its hypothesized impact, because one must also account for environmental, biological, and behavioral changes that may also affect the outcome. This issue, which is discussed in more detail in the next section, is frustrating because the impacts of health and nutrition programs take time to manifest and are often spread over individuals' lifetimes.

This problem of time lags warrants both a conceptual and a practical distinction between two types of programs. The first type is a program aimed at enhancing the stock of human capital by an intervention during some critical period of human physical and mental development. Programs involving mothers and children are of this type. The second type is a program aimed at increasing the efficiency of a flow of services from an existing stock of human capital. Programs involving adult workers are of this type (8).

The first type of program has long-run outcomes that are often not practical to measure because of the long time lag. Consequently, one must resort to proxies for measurements. The second type has shorter-run objectives, primarily increasing productivity and reducing absenteeism, that are immediately observable for adult workers. Termination of a program of the second type should end its effect and thereby provide another means of testing impact. Thus, selecting appropriate outcome measures is more problematic for the first, and more common, type of program. Consequently, this discussion focuses on outcome measures for the more general type of program with long-run impact.

In evaluating a potential outcome variable one should consider these questions:

- a. How does the variable relate conceptually to the ultimate outcome, or any component thereof, and to the program under study?
- b. How does it relate statistically to the ultimate outcome variable, or any component thereof, and to the program under study?

- c. How reliably can it be measured?
- d. What complementary data are needed?
- e. What are the costs of obtaining and using those data?

The first question is important because it relates an observed variable to the conceptual framework and to the specific hypotheses to be tested. The other questions bear largely on the statistical aspects of potential data for testing program impact. When all these questions are considered, trade-offs among particular types, or categories, of variables may appear.

We can classify outcome variables by three categories: inputs, intermediate outcomes, and ultimate outcomes. The use of program inputs as proxy measures of program impact is common. For example, the impact of a school feeding program is estimated by the amount of calories and protein the program delivers to the target population. This approach has the merit of being directly related to the program, and also is probably least costly since it is integrated with the program. However, it also tends to be the most presumptive since it may depend on hypotheses yet to be tested about relationships between inputs and eventual impact, and it may ignore program-induced behavioral changes beyond program control. This last issue depends critically on the delivery method; an income transfer to the household is easier for the household to divert from program objectives than, say, a directly administered vaccine.

Intermediate outcomes can be measured by a variety of variables: child morbidity, intellectual development, school achievement, and anthropometric measures including birthweight. These measures apply largely to children because theory and evidence suggest that they predict, and thus approximate, eventual health outcomes (9). Although they are not ultimate outcomes, such intermediate variables are "outputs" and in most cases they approximate ultimate outcomes better than program inputs do. Birthweight, for example, predicts relatively well a child's physical growth, at least during the first years and can be used as an outcome measure for maternal care programs (10). A child's physical growth and morbidity at an early age may indicate his future morbidity, survival probability, and productivity. A child's intellectual development, which to a degree may be nutritionally determined, is believed to be manifest eventually in his productivity and wages, primarily through mental development and school achievement (11-13).

These intermediate outcome variables raise the problem of lagged program impact since basically they are manifest over time.

Hence, linking the intermediate outcomes to the program inputs is even more complicated than using inputs as proxies for outcomes. Complementary data on non-program variables that affect the outcome may become critical for identifying the impact of the program. Consequently, collecting data that relate to the intermediate outcome variables requires more elaborate data collection instruments and statistical tools than when program inputs serve to measure outcomes.

Ultimate outcome variables involve issues similar to those of the intermediate outcomes, but they are more difficult since they pertain to full lifetimes. Outcomes and related variables can be measured on the basis of individuals, households, and communities. The choice of the measurement unit depends on the specific measurement objectives. Policy-makers and program administrators are eventually interested in variables that summarize their efforts on a community level. Students of household behavior are also interested in understanding and explaining the distribution of outcomes in the household or the community. Or, they may seek to understand why identical program inputs have a varying impact across individuals and households of different characteristics.

While critical for identifying the circumstances under which programs are beneficial, differences in impact may be concealed when we aggregate or, at times, disaggregate data. This possibility must be recognized when we define the unit of measurement. One cannot always distinguish household variables from variables pertaining to individuals. At times a variable can be based on a particular household member; at other times, it can be based on a few members. The definition of a household variable based on more than one household member may be complicated because of the low incidence of health-related events at the household level. Identification of the target population is a key criterion for the selection and definition of a household variable. If the target group consists of, say, mothers or potential mothers and we are interested in their nutritional status, then the observation is the mother -- an individual, as well as the household because in most societies we observe one mother per household. This household then can be described by other common household variables such as religion, income, size, or location, etc. The same reasoning applies when the target group consists of children of given age and sex.

A problem usually arises when one must aggregate within households or other units that share the same socio-economic endowments. This problem appears particularly acute in measuring some intermediate nutritional outcomes. For example, when the nutritional status of all school-age children in a household is of interest,

one needs to define a summary variable summarizing the nutritional status of these children in that household. The problem is that the number of children as well as age and sex distributions vary across households. This problem can be handled in various ways; however, while aggregating within the household, one must consider the possibility that not all children are treated equally in a given household.

The same problem applies to other household variables. For example, two mothers or two family units may live in one household and share a common income. Splitting such a group's income between the two family units, and subsequently treating them as separate units for statistical purposes, may be erroneous if behavioral patterns in extended households differ from the patterns in nuclear households. In most cases, when data are treated by averages, a few critical behavioral issues are assumed: that individuals are not discriminated against, and that the behavior of aggregated social units is the sum of the behavior of some other individual units. Before aggregating or disaggregating data, even at the household level, one must see whether these assumptions lead to different predictions.

AN ECONOMETRIC APPROACH

We turn now, before discussing program measurement, to outline some basic features of econometrics that draw on mathematics, statistics, and economic theory to delineate economic relationships empirically. The development of econometrics has been largely influenced by economists' traditional use of non-experimental data to test hypotheses and to verify economic relationships. 8/

Economic theory attempts to describe the nature of particular causal relationships, or "structures", by identifying the parameters encompassed by them, and by indicating the particular functional relationships among those parameters. As indicated previously, these relationships attempt to account only for what is believed to be the systematic part of man's economic behavior. In reality this behavior also has random elements and may be determined by parameters that we fail either to identify and observe, or to measure accurately. Consequently, the econometric equivalent of

8/ The term "econometric approach" used here is substantially the same as the "structural equations approach" (14). That is, it is a general application of mathematical statistics not necessarily restricted to "economic" variables. The reader familiar with this approach may bypass this section.

TABLE 1. REGRESSION COEFFICIENTS, CARBOHYDRATES, AND VITAMIN A AS DEPENDENT VARIABLES

(*t* statistics in parentheses)

Independent Variables	Dependent Variables (Nutrient)	
	Vitamin A * I. U.	Carbohydrates * g.
<u>Family Income*</u>	0.1542 (6.9459)	0.0085 (0.6250)
<u>Family Size*</u>	0.5130 (4.1073)	0.9135 (11.8791)
<u>North: Urban Elementary School (Intercept)</u>	10.7470	7.8833
<u>Region (South)**</u>	-0.1310 (-4.6454)	0.0579 (3.3276)
<u>Urbanization</u>		
<u>Non-Farm**</u>	0.0229 (0.5313)	0.2657 (9.9887)
<u>Rural-Farm**</u>	-0.0030 (-0.0864)	0.1489 (6.9906)
<u>Education</u>		
<u>High School**</u>	0.1852 (4.4095)	0.0486 (1.5837)
<u>College**</u>	0.1279 (3.7289)	0.0308 (1.4597)
<u>Adjusted R²</u>	0.743	0.920

* The logarithms of the variable values were introduced in the estimated equation.

** Denotes the use of a "dummy" variable which takes the value of 1 when an event occurs and the value of 0 otherwise.

the relationships suggested by theory include a "disturbance" or error term added to the "systematic" part of a particular relationship. This term summarizes the random, omitted, and unidentified or inaccurately measured elements of man's behavior. For example, if a particular simple economic relationship is characterized by

$$Y = f(X),$$

its econometric equivalent is

$$Y_i = f(X_i) + v_i \quad (i = 1 \dots n \text{ observations})$$

where v_i is the disturbance or error term.

Thus, econometric relationships also are *causal* relationships and are stated with one or more equations, depending on the underlying structure, each having a single dependent, or "outcome", variable and one or more independent variables (15); also see Annex, Sections A and B. Estimating such equations involves obtaining, usually by means of "statistical control", at least unbiased estimates of the particular effect on the outcome variable of each independent variable. ^{9/} Such estimates are possible when there is no, or only a "small", correlation between the error term and any of the right-hand variables, and between any two of these variables.

Given a particular economic model, at least three considerations are pertinent in specifying and interpreting econometric relationships: (a) choosing explicit functional relationships; (b) controlling for variables not suggested by theory; and (c) dealing with different and, therefore, competing hypotheses consistent with a particular estimate. The specification of a particular functional relationship can be based on common sense, theoretical and empirical knowledge, and experimentation. For example, we expect caloric consumption normally to level off as income rises, because of saturation and also because of substitution away from calories and carbohydrates to more "luxurious" proteins. Hence, the specifications of a functional relationship between caloric consumption and income should allow for a nonlinear relationship.

The two estimated equations shown in Table I are a simple example of econometric equations; the effects of certain household variables on household consumption of vitamin A and carbohydrates

^{9/} The common estimation procedure is regression analysis, which is also a useful descriptive tool. For biological applications of regression analysis, see for example (16).

are estimated. ^{10/} These estimates are based on a relationship like that shown in the Annex, Section B. This relationship allows for the expected nonlinear effect on consumption of nutrients by employing the logarithms of intakes of carbohydrates and vitamin A, and the logarithm of household income. This functional relationship was also chosen because it allows for a comparison between the sensitivity of the intakes of carbohydrates to income, and the intakes of vitamin A to income, regardless of the different units by which the two nutrients are measured, grams or international units. Such a relationship may impose various restrictions on the estimates; for example, the relationship specified presumes no consumption when the household has no current income. Nevertheless, such an unrealistic presumption and other restrictions do not outweigh the advantages of using this particular relationship. Another example of choosing functional relationships involves measurement of child growth by weight and height. Again, nonlinearity concerning age is appropriate here. The estimates shown in Tables 2 and 3 use a quadratic functional relationship that allows for this nonlinearity. These examples show how particular functional relationships are chosen on the basis of prior knowledge as well as practical considerations.

Controlling for variables not originally suggested by theory may be a useful means for improving the precision of the estimates, standardization, and adding information. ^{11/} An economic model of behavior may ignore biological factors by assuming them constant. For example, a model that deals with parents' choices of their children's diets and, therefore, child growth (18) might assume that the analysis is confined to a hypothetical age and sex group and, therefore, disregard the age and sex variables in explaining variations in child growth. However, for an econometric analysis based on a relatively small sample of children of both sexes and across age groups, one should control for age and sex as illustrated

^{10/} This table is drawn from (17) where an attempt is made to use economic theory and econometrics to predict and measure the effects of household characteristics on its diet. The coefficients on the logarithmic variables show the percentage change in household consumption of the particular nutrient due to a given percentage change in these variables. The coefficients on the dummy variables show the same but compared with variables which were "left out", or included in the intercept term. These estimates and others are used here for illustrative purposes only.

^{11/} Note that inclusion of omitted variables, which are not perfectly correlated with other independent variables, will increase the multiple correlation coefficient, or "explained variance" of the dependent variable. This increase should not be a goal by itself.

in Table 2, Equation 2. Otherwise, the relationship between child weight and age will be entirely approximated by the correlation between diet and weight across children of different age groups. 12/ This means of control is an alternative to various standardization procedures, like weight for age, used by nutritionists. It is also informative because it depicts growth curves directly across age and sex groups -- controlling for other effects -- and can be based on relatively small samples.

The problem of dealing with competing hypotheses is major. The first and basic issue to address is whether the line of causality implied by a particular relationship is correct. For example, in Table 2, Equations 2 and 3, the effect on children's weight of caloric intakes is estimated. However, the case can be made that the causality also operates the other way: heavier children consume more calories *ceteris paribus*. In a case like this it is reasonable to assume a structure where child weight and caloric consumption are codetermined. Therefore, one must use appropriate estimation techniques to allow for this codetermination or simultaneity. The estimates reported in these equations exemplify a case like this. The estimated effect of caloric consumption on weight in the third equation is based on a procedure that accounts also for the effect of weight on caloric consumption. This estimated effect differs from the estimate in the second equation where the estimation procedure does not account for simultaneity. 13/

The second issue is to ascertain that the variables indeed measure (or approximate) what they are meant to measure. In the context of nutrition interventions, a good example is the behavioral change in food consumption of families under observation due to presence of an interviewer. Thus, the variable that was supposed to measure the effect of program inputs measures instead another

12/ This procedure can be elaborated to account for interaction between the age or sex variables and other variables; that is, when, for example, a given diet has a different effect on children of different age and sex.

13/ Two or more variables are simultaneously determined when they are outcomes or endogenous variables belonging to a structure where one of these variables affects the other in some relationships and viceversa in others. Failing to account for this codetermination while estimating one particular structural relationship may result in a "simultaneity bias" which means that the estimated effect of one endogenous variable on the other may be under- or overstated. On how to deal statistically with such cases, see (15); for specific application in a nutrition study from which this table is drawn, see (19).

TABLE 2. REGRESSION COEFFICIENTS, CHILD'S WEIGHT IN KG AS A DEPENDENT VARIABLE, ORDINARY LEAST SQUARES (OLS) AND TWO-STAGE LEAST SQUARES (TSLS) ESTIMATES

(*t* statistics in parentheses)

Equation No.	Type of Estimation	Intercept	"Biological" Variables				
			Age	Age ²	Sex Male*	Age Sex	(Age Sex) ²
1	(OLS)	-0.63771	0.609 (6.628)	-0.008 (-4.736)	7.533 (5.526)	-0.563 (-4.280)	0.011 (3.708)
2	(OLS)	3.173	0.245 (3.559)	-0.019 (-1.261)	0.708 (2.994)		
3	(TSLS)	1.788	0.112 (1.246)	-0.0001 (-0.077)	0.169 (0.272)		

(Equations continued from above)	Economic Variables				
	Occupation			Diet	
	Land Owner*	Agri. Labor*	Civil Servant*	Calories	R ²
1	0.860 (2.559)	0.038 (0.110)	0.745 (1.851)		0.61
2				0.001 (4.125)	0.55
3				0.004 (3.235)	

* Dummy Variables

behavioral change induced by the program, thereby founding the measurement of the program impact. Similarly, the estimated effects of the sex variable on children's nutritional status may measure parents' behavioral discrimination between sons and daughters, as well as genetic differences between sexes. In a case like this, it may be impossible to discriminate statistically between the two competing hypotheses. ^{14/} Such cases put an added burden on the scientist beyond getting unbiased estimates; he must be careful in interpreting his findings in light of a variety of behavioral and other relationships that can produce particular statistical results.

Some cases of competing hypotheses are more predictable and statistically manageable than others. These cases can be generally characterized as those in which one has some notion or knowledge about correlations between only two independent variables. For example, the United States food stamp program is designed for low income groups. Consequently, a variable representing household participation in the program is expected to be negatively and highly correlated with the income variable in an analysis of this program's impact across income groups. Therefore, statistical discrimination between the effect of the program as opposed to the effect of income on a particular outcome variable may present a significant problem. A correlation between the disturbance term and one or more independent variables may have similar consequences.

Econometrics is concerned with the problems briefly described here and other related ones. The solutions offered are geared, however, to existing data and, consequently, are largely based on statistical control.

IMPACT MEASUREMENT: ECONOMETRICS AND SOCIAL EXPERIMENTS

An illustration based on one of the major "experiments" in nutrition can be helpful to the more general discussion that follows about the use of econometrics in measurement of program impact and the relationship between econometrics and social experiments.

The estimated regression coefficients shown in Table III are based on the Narangwal experimental project (21). The first equation -- which is equivalent to comparing the mean outcomes of an experiment -- shows the estimated effects on child weight of various

^{14/} For a technical discussion on related issues in the context of a nutrition project, see (20).

TABLE 3. REGRESSION COEFFICIENTS, CHILD'S WEIGHT IN KG AS
A DEPENDENT VARIABLE

(*t* statistics in parentheses)

Equation No.	<u>"Biological" Variables</u>			<u>Socio-Economic Status</u>	
	<u>Intercept</u>	<u>Age</u>	<u>(Age)²</u>	<u>Sex Male</u>	<u>Higher Caste (JAT)*</u>
1	7.96897				
2	3.18148	0.00967 (10.81462)	-0.00001 (-4.59184)	0.51932 (4.00037)	0.52166 (3.85227)

(Equations conti- nued from above)	<u>Types of Intervention</u>			<u>R²</u>
	<u>Nutrition Supplement*</u>	<u>Medical Care*</u>	<u>Nut/Supp. + Med. Care *</u>	
1	-1.12432 (-2.45703)	-0.08715 (-0.18166)	0.27777 (0.69498)	.064
2	0.77071 (2.76640)	0.22200 (0.80560)	0.76378 (3.34425)	.707

* Dummy Variables

interventions, without controlling for other variables that might affect the outcome (see Annex, Section E). The second equation shows the same, but with statistical control for age, sex, and socio-economic status approximated by caste. The first equation shows that nutritional supplementation has a negative and statistically significant effect on child weight, while medical care and a combination of medical care and nutritional supplementation had no effect. Once we control for the other variables, intervention by nutritional supplementation and by nutritional supplementation combined with medical care show a positive and statistically significant effect on child weight.

The Narangwal experiment was not a "true" experimental design, but for that matter, it probably could never be. ^{15/} Moreover, it represents well the constraints of the kind most field programs face. Treatments were applied to entire villages because random assignment of the treatments within villages was socially and practically inconceivable. Randomization on the basis of a population of villages was impractical because it would inflate the program to proportions beyond its financial and logistic means. ^{16/} Two issues had to be considered before estimating program impact. First, although similar, the villages may have been different also in aspects other than the treatments. Second, there was scope for behavioral "self selection"; different people could, by choice, avail themselves differentially of the services, and for that matter, could benefit differentially from given services.

Indeed, the combination of these factors is evident in the results presented here. The first equation shows merely a negative correlation between child weight and nutritional supplementation. This is consistent with either or a combination of the following

^{15/} "Experimental design" refers to use of a "control" group and "treatment" group (or groups) selected entirely by a random process. In the purest and simplest type of randomization or experimental design, as opposed to an econometric approach, one is not concerned with identifying and specifying particular causal and structural relationships and with estimating their various parameters. Full discussion of experimental design is beyond the scope of this paper. The discussion here is based on (22-24).

^{16/} Almost none of the major other field experiments, like those undertaken by INCAP, meets the requirements of experimental design. The same is true, for that matter, in the field of family planning (25). It appears that researchers attempted, for practical and ethical considerations, to approach a situation where they have "matching groups".

hypotheses. First, the village where children received nutritional supplementation could have had children with a lower average weight, possibly because of nutritional reasons as well as the children's age and sex composition. Second, higher caste and better caring mothers, who have heavier children, could have used the nutrition program more, and more efficiently, than others. ^{17/} Once we control for these possibilities, by adding the other variables to the estimated relationship, we get better estimates of the interventions' impact in the short run as measured by child weight.

This example can be useful in outlining a general (and pragmatic) approach as well as specific criteria for assessing statistical requirements for evaluation research. However, we must first agree on the following points: (a) on pure statistical grounds, (randomized) experimental designs based on large samples are superior to any other approach for "netting out" impact; (b) quantitative evaluation of programs must be, however, socially efficient; and (c) to meet the efficiency requirement we must always weigh the value of the additional (marginal) gain in statistical evidence and accuracy versus the marginal costs involved. The last point is a general efficiency criterion that may dictate the use of experimental design, econometrics, and a combination of the two. The combined use of econometrics and experimentation is emerging in economic studies, for example, the New Jersey Graduated Work Incentive Experiment (26).

Returning to the Narangwal example with the above points in mind, we must first address what it represents, and then consider what the results mean in terms of the relative efficiency of social experiments. With respect to the first question, it represents what economists call a "revealed preference": the best that could, and probably can, be done, given objectives and constraints. This does not mean that we cannot have better field experiments. It merely puts a question mark on their economic and political feasibility.

On the issue of the relative efficiency of social experiments, the results obtained by means of statistical control are most likely biased, or lack "internal validity", at least because of the nonzero correlations between any two of the right-hand variables. However, these biases do not appear serious, and the results are fairly "reasonable". ^{18/} That is, it is highly probable that a

^{17/} Other possibilities not discussed here are also plausible.

^{18/} The results indicate that using the short term outcome, child weight, nutritional supplementation had a measurable effect. Due to this particular intervention child weight in the sample used

full-fledged (and costlier) experiment might not improve the results substantially. While this is a testable proposition, which warrants an experiment, the author is willing to conclude that to determine the impacts of the Narangwal project, the mix of "semi-experimentation" and a subsequent econometric approach prove sufficient.

To outline an econometrician's general approach to data requirements for impact measurement, let us consider three interventions aimed at increasing caloric consumption: (a) a national food subsidy program to reduce the prices of cereals to all consumers; (b) a national food stamp program for everyone; and (c) a national food stamp program with randomly selected participants. Assume that the three programs will give participants identical price reductions. ^{19/} Program (c) can be evaluated simply and unambiguously just by collecting caloric consumption data after the fact, if we are content merely to answer two questions: Did the program increase caloric intakes? If so, what was the average increase per family? The answers to these questions could be unequivocal, assuming we gather accurate data and we draw a large enough sample. But this approach would tell us virtually nothing about why some families respond to the program more or less than the average or in the case of program failure, why there should have been no significant impact. In more general terms, the experimental program (c) lacks "external validity". Moreover, apart from cost considerations, the administration of this program poses social and other hazards that may jeopardize the validity of the experiment.

Programs like (a) and (b) may be more realistic and are more common than an experimental program like (c). Taking this reality as a constraint, we stress how one can measure impact using an econometrician's perspective. Although programs (a) and (b) cause identical reductions in prices, there is a fundamental difference between them. Under the first program, households can buy the cereals only at the subsidized price. Under the second, households can exercise choice in buying cereals with or without food stamps,

here increased by .77 kg., or 9.7%, at the sample mean. The highest zero-order correlation between two of the right-hand variables was .34. It should be noted that a more careful econometric treatment of the data presented here may improve the results.

^{19/} Other differences between the programs are ignored. For a discussion of the different implications of programs, like (a) and (b), see (27).

and they can even sell the stamps. ^{20/} This difference between the programs is also of key statistical significance. For the first programs, all we know is that it started at some particular point in time and existed to a subsequent point. But for the second, additional important information is potentially available: whether or not, and possibly to what extent, a household used the program stamps for buying cereals.

To determine impact in the first instance, panel data are necessary. Pre-program, or "baseline" data, should include the variables believed, on theoretical and other grounds, to affect the outcome. These data serve to estimate the basic relationship between, say, caloric consumption and its nonprogram determinants. Once this relationship is estimated, we may also be able to estimate the added program effect (c.f. Annex, Section D). For example, suppose the food subsidy is introduced in an agricultural economy and reduces prices of cereals. However, if farmers' incomes fall during the program period because of bad weather, for example, the farmers may consume less cereals even at the lower prices. Hence, such a fall in incomes may offset any positive effect of the program. Therefore, by accounting for changes in incomes (or even weather conditions), one can estimate program impact, provided no event significantly affecting the program outcome occurs simultaneously with the program in a way to confound our inferences.

In the second instance, the national food stamp program (Program B) baseline data may not be needed if we can analyze statistically the variance in households' utilization of the program. This possibility depends on the relationship between program utilization and other variables. At least from a behavioral standpoint, it is almost inconceivable that households of different characteristics will use, and even benefit from, the program identically. Therefore, it may be possible to study the determinants of program utilization as well as the interaction between program inputs and household characteristics. That is, what are the effects on utilization of various income and education levels? How do households of different levels of income and education benefit from given program inputs? If consistent answers should emerge, it might be relatively straight forward to infer about the impact of the program itself (c.f. Annex, Section C).

^{20/} Although the food stamps are available to all, one must consider the time and trouble of getting them. For example, a mother with five small children, living at a distance from where food stamps are sold, may find it less attractive to obtain those stamps than a mother of five grown children living nearby that selling station, *ceteris paribus*. See Annex, Section C.

A program involving a target population identified by socio-economic characteristics, geographical location, or nutrition status can be regarded as a special case. Such a program differs from those already discussed in that it is for a population identified by one or more of the non-program variables affecting the outcome. Hence, there is a built-in correlation between the measure of participation in the program and some other determinant(s) of the outcome. Most policy programs fall into this category. Impact measurement in such cases must be confined to the target population and involves essentially identical considerations to those relating to the non-experimental programs discussed above. When within the target population, there is no (or not enough) variation in the program (input or utilization) variable, panel data will be required to measure impact. For example, impact measurement in a particular intervention village of the Narangwal project is not possible if all that is known is that that particular village was exposed to that intervention. However, when there is sufficient variance in the program variable within the target population, this variance can be used to measure impact. For example, the United States food stamp program is designed for low-income families. Within that population, however, some families do not use the program at all and other families use the program to different degrees. This variation may permit estimating program impact.

Within the perspective taken in the preceding paragraphs, an experimental approach is warranted when it is less costly, or when there is reason to believe that, given the available statistical techniques, panel data are insufficient to measure impact. Two basic situations may undermine the use of panel data. The first is when the underlying structural relationships are believed to change over time. Such a change may prohibit estimating the added effect of the program. The second situation is one where there is reason to believe that one or more of the non-program determinants of the outcome will change during the monitoring period in a way that will confound any inference. Hence, experimental design may be used as a precautionary measure against foreseen and unforeseen circumstances. However, taking the view that social experiments are costly, financially and otherwise, we must look for ways (a) to reduce their costs and (b) to maximize their usefulness.

To reduce costs, we must protect the measurement procedure against uncertainty rather than against ignorance because protection against ignorance can be unnecessarily more costly. To elaborate, uncertainty is defined here as a situation where one has good reason to believe that some ~~known~~ non-program determinants of the outcome may change significantly during the program and potentially have a confounding effect. For example, when an outbreak of diarrhea may confound the measurement of the impact of a

child-feeding program, implementation of experimental child-feeding programs in two or more areas will lessen the risk that diarrheal disease will confound the results and obscure a possible treatment effect. Ignorance is defined here as a situation when one does not know the determinants of the outcome variable. In this case, one may wish to have an experiment by a random assignment of the intervention to "protect" the evaluation from all possible confounding effects. Unfortunately, this may be nearly impossible and unnecessarily costly.

When considering the use of experimental design, the economic theory of the household and econometrics may help to reduce the "costs of ignorance" by suggesting which non-program variables might confound the evaluation and need to be controlled for. Alternatively, prior considerations should help to stratify and reduce the evaluation sample. ^{21/} The first step is to understand, possibly by some modeling, the behavioral and other determinants of a particular outcome. This effort should be coupled with anthropological observations as well as a pilot survey. Once key non-program determinants of the outcome have been identified, the number of variables for which the experiment has to be controlled may become manageable. The second step is to protect against uncertainty by limiting the control variables to those that may change during the life of the program. This step may further reduce the cost of the evaluation. Obviously, in reality the researcher confronts both uncertainty as well as ignorance. However, given the costs of evaluation research based on experimental design, the benefits from careful *a priori* modeling may be substantial.

When experimental design is selected, an econometric approach to the data can help maximize its usefulness. As already mentioned, from a statistical viewpoint, impact measurement in a well-designed experiment can be satisfied just by comparing the means of the outcome variables between the treatment group(s) and the control group(s). This approach, however, reduces the acceptability and the universality of the results and has serious shortcomings in experimental contexts where policies are tested for future and wider applications. To increase the generality, or "external validity", of an evaluation based on experimental design, the researcher should include other variables or controls that affect the outcome measure and apply econometric techniques to the data (cf Annex, Section E). In this way, one can test some basic behavioral hypotheses and perhaps show explicitly how a program relates to its environment.

^{21/} This approach comes close to a "quasi-experimental" design.

CONCLUSION

The basic postulate of this paper is that although health related interventions aim to increase household and individual welfare, they may be inconsistent with the household's and individual's own objectives and opportunities. Where households can exercise choices *vis-a-vis* the program, studying these objectives and opportunities and relating them to particular interventions must be the first steps in understanding and predicting a program's potential impact. While analyzing household behavior is no mean task, it is nevertheless essential. By relating the concept of expected earnings to age-specific productivity, morbidity, and mortality probabilities, the economic theory of the household offers a conceptual framework in which to integrate relevant aspects of household behavior with the basic outcomes of health programs. Thereby, this theory can suggest testable hypotheses concerning program impact, and can help to identify various outcome measures for programs.

Econometrics stresses the empirical aspects of the causal behavioral relationships outlined by the economic theory and, given the social and financial costs of social experiments, can provide a useful and efficient framework for measuring program impact. An econometric approach can, at times, substitute for experimental design, and at other times, complement it, depending on financial and other costs as well as statistical considerations. When experimental design is warranted on statistical grounds because statistical control techniques are insufficient to measure impact, the economic theory of the household may help to minimize the cost of the experiment by suggesting which variables should be controlled. Furthermore, once data from experimental design are available, the use of an econometric approach to explore those data can increase the credibility and generality of the results.

Thus, the combination of theoretical and other *a priori* considerations with econometric estimation techniques and experiments can prove the most efficient way to measure program impact.

ANNEX

A. A simple economic relationship -- say between a household's caloric consumption (Y) and a vector X of household characteristics including, for example, household income -- can be stated by

$$Y = f(X). \quad [1]$$

This is an "exact" relationship in that it does not account for random elements in human behavior. It is also a partial relationship in the sense that it may "ignore" other lines of causality;

a more comprehensive underlying model may also deal, for example, with the effects of caloric consumption on household income. Given a relationship like [1] economic theory is usually concerned with the nature of the changes in Y due to changes in X, within a relevant range.

B. Assuming that X represents just income and that Y and X are observable, the econometric equivalent of relationship [1] can, for example, be

$$\log_e (Y_i) = a + b \log_e (X_i) + v_i, \quad (i = 1 \text{ --- } n), \quad [2]$$

where the index i denotes one of n observations, and v_i is a disturbance, or error, term which is usually assumed to be random, to have a normal distribution with an expected value of zero and a constant variance, and to be uncorrelated with other independent (right-hand) variables and with similar disturbance terms across observations. The term "a" is a shift parameter or "intercept", indicating levels of caloric consumption that are independent of X and v. The term "b" indicates, in this case, the change in log (Y) due to a unit change in log (X). Alternatively, "b" indicates the percent change in Y due to a given percent change in X, or the "elasticity" of Y with respect to X. 22/ Assuming no errors in the specification of [2], obtaining an unbiased estimate of "b" is possible when X_i is not correlated with v_i . 23/

C. The following underlying structure is generally assumed here for discussing measurement of the impact of a program. First, in a given environment where such a program has been implemented, an outcome (Y) is a function of the household's utilization of the program (U) and particular subset of household characteristics (X_1); that is,

$$Y = f (U, X_1) \quad , \quad [3]$$

where U and X_1 may be dependent; that is, for biological and behavioral reasons, particular levels of program utilization may have different effects in households of different characteristics; $\frac{\partial^2 Y}{\partial U \partial X_1} \neq 0$.

Second, the utilization of the program is a function of the supply of program inputs (P) indicating the intensity by which the

22/ Some other properties of this double-logarithmic function are discussed in Section IV of the text.

23/ When X_i denotes a vector of variables, any two of those variables should also be uncorrelated.

program reaches different households, and of a subset of household characteristics (X_2);

$$U = g(P, X_2) \quad , \quad [4]$$

where X_1 and X_2 are not necessarily mutually exclusive subsets at least conceptually.

This structure can be reduced by substituting (4) into (3), to

$$Y = f[g(P, X_2), X_1] = h(P, X_3), \quad [5]$$

where X_3 is a set of household characteristics combining X_1 and X_2 .

Estimating relationship [5] may be sufficient to measure program impact. This alone is insufficient, however, for understanding this impact and, thus, for evaluating the program. In particular, such a relationship does not reveal whether variations in impact of given program inputs are due to different levels of utilization, to differences in impact of given inputs on households of different characteristics, or both. The problem can be stated as the researcher's inability to test various hypotheses concerning X_1 and X_2 . This problem can be solved by estimating relationship [4], which may be crucial for designing a cost-effective program because this relationship identifies the users and levels of utilization of a particular program. 24/

D. A simple approach to panel data pertaining to a program is to pool baseline and program data, and estimate an explicit function of

$$Y_{it} = g(t, X_{it}, P_i, v_{it}), \quad (i = 1 \dots n, \quad t = 0, 1) \quad [6]$$

where: t represents the time period a particular observation was obtained; 25/ Y_{i0} and X_{i0} represent baseline or preintervention values; $Y_{i1} = Y_{i0} + dY_i$ and $X_{i1} = X_{i0} + dX_i$ represent postintervention.

24/ Although conceptually different, P and U are often interchanged and relationship [3] is estimated instead of [5]. In such a case the researcher should be aware that he may attribute undue failure or success to program services. That is, the program may succeed or fail because of particular household characteristics; some programs may succeed in one setting and fail in another.

25/ $t = 0$ for preintervention and $t = 1$ for during or postintervention. This illustration can be generalized to include more than two time periods.

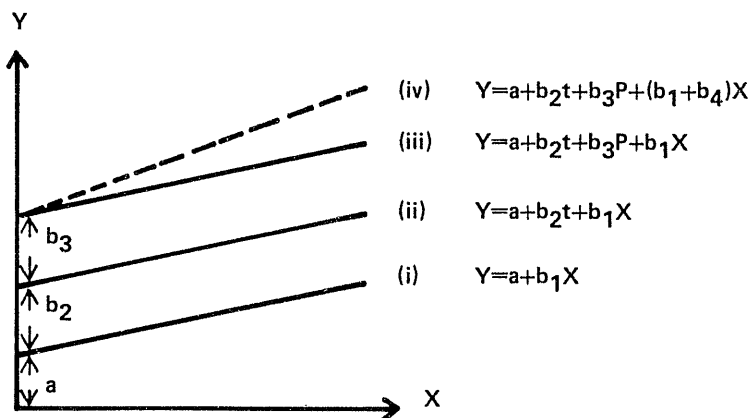


Fig. II. General Relationship Between Income and Caloric Consumption.

tion values; and P_i indicates program inputs. ^{26/} Assuming, without much loss of generality, that the relationship between caloric intakes and the other variables is linear within the relevant range, and that changes in income bring about identical changes in consumption over time, one can estimate this equation as illustrated in Figure II.

$$Y_{it} = a + b_1 X_{it} + b_2 t + b_3 P_i + b_4 P_i X_{it} + v_{it} \quad (7)$$

First, no interaction is assumed between the program (inputs) and household characteristics, $b_4 = 0$; that is, program impact is independent of the level of household income. Curve (i) in Figure II indicates the general relationship between income and caloric consumption, or the pre-program relationship. Curve (ii) indicates the added effect of time, measured by b_2 that is presumed to be independent of the other effects. For example, it can indicate the effect of child growth on caloric consumption. Curve (iii) depicts the added effect of the intervention measured by b_3 .

Second, we assume that Y measures child weight, for given age and sex, and that there is interaction between X and P , or $b_4 \neq 0$;

^{26/} Alternative, $P_i = 1$ for participation in, or prevalence of, the program and $P_i = 0$ otherwise.

that is, for a given level of program utilization the impact of the program, measured by gain in child weight, differs across households of different levels of income. For example, children of better-to-do families may be less exposed to diarrhea than children of poor families. Consequently, the children of relatively richer families may gain more in weight from a given school feeding program, *ceteris paribus*. This particular effect is depicted by line (iv) of the figure.

Estimating program impact by using panel data is not feasible when "p" and "t" are statistically indistinguishable; that is, when some other events potentially affecting the outcome concur with the program.

E. Suppose that the sample underlying relationship [7] and Figure II is drawn from an experimental design which consists of n observations, m in the control group and $(n-m)$ in the experimental or program group. That is, $P_i = 0$ for $i = 1 \dots m$, and $P_i = 1$ for $i = m+1 \dots n$. To measure and test for intervention impact, the experimental design advocates measuring the difference $dY_e - dY_c$ and testing for its statistical significance, where $dY_c = dY_i/m$ is the mean change in the outcome in the control group, and dY_e is the corresponding statistic for the experimental group. This difference will reflect correctly and fully program impact, only if the experimental design achieves complete randomization (or perfect matching) of all other potential effects; that is P_i does not correlate with any other determinant of the observed outcome.

To achieve the same goal an econometrician would estimate relationship [7], by either

$$Y_{it} = G + b_3'' P_i + V_{it} \quad [8]$$

$$\text{or } dY_i = E + b_3'' P_i + W_i \quad [9]$$

where W_i is a disturbance term similar to V_i . It can be shown that

$$dY_e - dY_c = b_3 = b_3' = b_3'' \quad [10]$$

when P_i does not correlate with any of the right-hand variables in equation [7]. Hence, while econometric and experimental approaches should yield identical statistical results because of the underlying statistical rationale, the econometric approach specifies particular structural and functional relationships and tries to account for all relevant determinants of the outcome.

ACKNOWLEDGEMENT

My colleagues A. A. Kielmann, M.D., and D. Coate, as well as participants in the PAHO conference for which this paper was prepared, should share in any credit for this presentation. However, neither they nor the World Bank should be held responsible for the views expressed here.

REFERENCES

1. World Bank. *Health Sector Policy Paper*. Washington, D. C.: World Bank, 1975.
2. Becker, G. S. The theory of allocation of time. *Economic Journal*, 75:493-517, 1965.
3. Lancaster, K. J. A new approach to consumer theory. *J. Political Econ.* 74:132-57, 1966.
4. Eckholm, E. and F. Record. *The Two Faces of Malnutrition*. Washington, D. C.: World Watch Institute, Paper No. 9.
5. Mushkin, S. Health as an investment. *J. Political Econ.* 70:129-157, 1962.
6. Grossman, M. *The Demand for Health: A Theoretical and Empirical Investigation*. New York and London: Columbia University Press for National Bureau of Economic Research, 1972.
7. Mincer, J. *Schooling, Experience and Earnings*. New York and London: Columbia University Press for National Bureau of Economic Research, 1974.
8. Basta, S. S. and A. Churchill. *Iron Deficiency Anemia and the Productivity of Adult Males in Indonesia*. Washington, D. C.: World Bank, Staff Working Paper No. 175, 1974.
9. Stoch, M. B. and P. M. Smythe. Fifteen-year development study on effects of severe undernutrition during infancy on subsequent physical growth and intellectual functioning. *Archives of Disease in Childhood*, 51:326-336, 1976.
10. Lechtig, A. Effect of improved nutrition during pregnancy and lactation on developmental retardation and infant mortality. In: P. L. White and N. Selvey (eds.) p. 117, *Proceedings of Western Hemisphere Nutrition Congress*, Massachusetts: Publishing Sciences Group, Inc., 1975.
11. Selowsky, M. A note on preschool-age investment in human capital in developing countries. *Economic Development and Cultural Change*, 24:707-720, 1976.
12. Selowsky, M. and L. Taylor. The economics of malnourished children: An example of disinvestment in human capital. *Economic Development and Cultural Change*, 22:17-36, 1973.
13. Grossman, M. and L. N. Edwards. *An Economic Analysis of Children's Health and Intellectual Development*. New York: National Bureau of Economic Research, Working Paper No. 180, 1977.

14. Duncan O. D. *Introduction to Structural Equation Models*. New York: Academic Press, Inc., 1975.
15. Johnson, J. *Econometric Methods*. Second Edition, New York: McGraw-Hill, 1972.
16. Snedecor, G. W. and W. G. Cochran. *Statistical Methods*. Sixth Edition, Ames: Iowa State University Press.
17. Chernichovsky, D. The demand for nutrition. Washington, D.C., World Bank, 1977 (Mimeograph).
18. Chernichovsky, D. and D. Coate. *The Choice of Diet for Young Children and Its Relation to Children's Growth*. New York: National Bureau of Economic Research, Working Paper No. 219, 1977.
19. Chernichovsky, D. and A. A. Kielmann. Socio-economic status, diet and preschool child growth in rural Punjab, India. Baltimore: The Johns Hopkins University Press, 1977. (Mimeograph).
20. Nasoetion, A. H. Spurious correlation as a result of constraints in randomization. *Research Methodology, Agricultural Development Teaching Forum Council*, 46: 1975.
21. Taylor, E. C., et al. Malnutrition, infection, growth and development: The Narangwal experience. Baltimore: The World Bank and the Johns Hopkins University Press, 1978. (Mimeograph).
22. Campbell, D. T. and J. C. Stanley. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand-McNally Publishing Company, 1966.
23. Ross, J. and P. Smith. Orthodox experimental designs. In: H. M. Blalock and A. B. Blalock (eds.) p. 339. *Methodology in Social Research*. New York: McGraw-Hill, 1968.
24. Bennet, C. A. and A. A. Lumsdaine. *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs*. New York: Academic Press, Inc., 1975.
25. Cuca, R. and C. Pierce. *Experiments in Family Planning*. Baltimore and London: The Johns Hopkins University Press for the World Bank, 1977.
26. United States Department of Health, Education and Welfare. *Summary Report: New Jersey Graduated Work Incentive Experiment*. Princeton: New Jersey, 1973.
27. Reutlinger, S. and M. Selowsky. *Malnutrition and Poverty*. Baltimore and London: The Johns Hopkins University Press for the World Bank, 1976.

COMMENTS

Sebastián Piñera, *Comisión Económica para América Latina (CEAL)*
Santiago, Chile

Chernichovsky's paper treats two conceptual matters and one methodological matter that have fundamental importance for evaluating the impacts of nutrition and related health programs.

The first conceptual matter is the potential contribution of the "New Home Economics" to understanding the determinants of program impact. The need to evaluate impacts of nutrition and health programs arises from the existence of resource scarcity in face of multiple uses. Therefore, efficiency requires selection of programs with major impacts on the variables one wishes to affect. Moreover, a useful evaluation not only should quantify the aggregate success level of an intervention, but also should locate and identify -- at each stage upon which impact depends -- the leading factors that tend to raise or lower the degree of success. Hence arises the importance and contribution of the "New Home Economics".

This approach regards the household as the basic socio-economic unit within which decisions are made about allocation of the household's resources. These decisions include its supply of labor and capital to factor markets as well as its consumption of goods and services. So long as the household is the primary decision-making unit, analyses of the determinants of household behavior seem fundamental for non-experimental statistical evaluations of nutrition and health programs impacts. In the context of this theory, household decisions are made so as to maximize a utility function (or, more fully, a "happiness" function) subject to certain constraints imposed by initial factor endowments, by technical production possibilities for self-consumed goods and services, and by the prevailing prices in different markets. In light of this focus, an intervention program will have an impact only to the extent that it affects the preferences (or the utility functions) of households, or if it affects some of the constraints that shape households' maximizing behavior. In other words, an intervention program will only be able to change the combination of signals, incentives, and stimuli upon which the family unit bases its decisions: income levels, prices, knowledge about the utility of certain goods, and so forth. But without a doubt, it is the household that determines impact by revising its maximizing behavior in light of new conditions introduced by the program. Recognition of this fact is of fundamental importance for correct evaluations.

Some programs may cause such drastic changes in incentives or signals that they leave little or no room for family decisions

that might affect impact. Examples of such programs are obligatory vaccination and obligatory or strongly promoted consumption of pills or contraceptives. However, even for these sorts of direct interventions, family members can change behavior so as to make substitutions that modify the intervention's impact. For example, women that get rewards or subsidies for using contraception can simply employ the new methods as substitutes for those used before, perhaps reducing their ages of marriage or the lengths of their lactation periods. The final result is that a program, even though it distributes many contraceptives, may have no impact upon fertility. Likewise, the parents of children that get free milk in school can reduce the amount of milk consumed at home so that their children's total milk intake remains constant. Once again, the program's impact on children's total milk consumption can be zero in spite of the direct milk distribution.

In less direct programs, family behavior and decisions are even more important in determining success. Examples are programs to improve poor families' nutritional status via income subsidies, food price subsidies, free distribution of food to families, and so forth. With such programs, impacts on the nutritional status of poor families, particularly for members like children and the elderly, depend critically upon family behavior. Income and food price subsidies may have minimal impacts on nutrition when marginal propensities to consume food and price elasticities of demand for food are low. By the same token, under certain conditions, the free distribution of food can be exactly equivalent to an income subsidy and can have an insignificant impact on the nutritional situation it is supposed to affect. Moreover, this sort of program can increase food consumption of certain groups within the nuclear family (adults and men) and either not affect or affect minimally the consumption of others (children and women). In this respect, the few studies of intrafamilial distribution of food consumption show the marginal propensity to consume food varies significantly across members of the family and types of food. In other words, the important thing for measuring program impact is not which resources or what quantity of the food in question was supplied by the program, but rather which of these resources or of this food were actually used by the target population. (One should not forge' that powdered milk supplied by a nutrition program may also be used to paint walls, to mark football fields, and so forth!).

In summary, to guarantee the success of an intervention, it is not enough to create a greater supply of the goods or services one seeks to encourage. As important as supply is, there is also a critical need to induce or generate a demand from the target population for these goods or services. Participation of the target population in all stages of the program can be an indispensable

or very important element for the success of this last objective. It can succeed through programs of education or incentives for the target population to participate.

In general, each intervention program tries to produce a pattern of consumption different from what the program's target households would have followed if an amount of money equivalent to the program's resources had simply been made available through income transfers. Implicit in this attitude is the notion that the households or people supposed to benefit from the program are not capable -- due to ignorance, lack of information, or whatever other reason -- of ordering correctly their wants or the distribution of these wants among different household members. The fact that the program, in order to be successful, cannot respect "consumer sovereignty" and must overcome the "erroneous" choice of wants, underlines the importance of the demand aspects of intervention programs.

The second crucially important conceptual feature Chernichovsky discusses is the distinction among inputs, intermediate outcomes, and ultimate outcomes as program impact measurements variables. In this regard, Reutlinger and Selowsky write:

"Just as the educational level of a population should ideally be measured by its educational achievements and not by its exposure to, or use of, educational inputs, so malnutrition should ideally be defined by its consequences, such as health status, rather than by nutrient intake. In practice, it is difficult to define objective indicators of consequences, and it is even more difficult to collect and interpret relevant data" (1).

It is exactly this difficulty in defining objective indicators of the intermediate and ultimate outcomes of intervention programs that has caused most evaluations of program impact to tend to be based upon program inputs more than upon intermediate and ultimate outcomes. In light of the discussion with respect to the role of household behavior as a determinant of program impact, the difficulties of basing program evaluation upon inputs shows up clearly. For this reason, Chernichovsky's distinction among these three types of measurement variables is extremely important and illustrates the advantage of investigating evaluation approaches based upon programs' intermediate or ultimate outcomes.

These sorts of approaches can avoid certain circularities incurred by approaches based upon inputs. For example, traditionally a person's nutritional status is measured by the number of

calories and other nutritive elements consumed daily. Minimum caloric consumption standards depend among other things upon a person's weight, which in turn depends upon this person's caloric intake -- thus introducing an element of circularity.

The third very important item treated by Chernichovsky is related to methodology. He refers to the advantages and disadvantages of the experimental (or quasi-experimental) methods and of the statistical control methods as alternatives for evaluating the impact of intervention programs. This point is particularly well treated by him. Because of the importance of this matter to the field of program evaluation, it seems to me helpful to repeat from the paper a phrase with which I agree fully:

"In a social setting experimental design may be more expensive -- financially, politically, and morally -- than measuring impact by means of statistical control".

The constant progress of econometrics and computation techniques, along with the constant decrease in computation costs, tend to make the method of statistical control even more attractive as an instrument for evaluation of intervention programs.

REFERENCES

1. Reutlinger, S. and M. Selowsky. *Malnutrition and Poverty*, p. 8, Washington, D.C.: World Bank, 1977.

GENERAL DISCUSSION

The Chairman emphasized the dual nature of Chernichovsky's paper. Its two main themes were (1) the household as a decision locus; and (2) the use of statistical control, particularly the methodology developed by econometricians, as an alternative to experimental evaluations.

Much of the discussion dealt with costs, particularly with the relative costs and complexities of experimental (or quasi-experimental) designs as opposed to evaluations via multivariate statistical control. One speaker emphasized his concern that expensive evaluations might take funds needed for operational activities and suggested that no more than three percent of total project costs be allowed for evaluations (except for projects clearly labeled as "research" or "pilots"). He noted that China probably has the largest nutrition program of any developing country, and yet it spends almost nothing on formal evaluation. He said the Chinese results are plain to see, in the form of healthy children, so he thought further "evaluation" there was

unnecessary. He recommended Latin American programs steer a middle course between the Chinese extreme of no evaluation and the North American extreme of highly sophisticated and costly evaluations. But several other speakers disagreed. One said the three percent target would sometimes be too high, other times too low, all depending upon the general conditions of health in the project area, the demands for and use of services, the human and financial resources available, and the organization and functioning of the local health system. He said the appropriate level of complexity and cost for any evaluation would also depend upon what the decision-makers actually want to know (and are willing to act upon) in any particular situation and upon what information the researchers and evaluators can actually provide. He also suggested there may be circumstances in developing countries where evaluation and research are *more* important than for similar projects in the industrialized countries, since the developing countries may not be able to afford enough "general" development investment (for higher income and education, for example, as opposed to specific nutrition and medical interventions) to raise overall nutrition and health levels. And good evaluations might help developing countries avoid funding certain very costly programs that would be likely to fail. The speaker said he knows several past examples where relatively unsophisticated evaluation techniques helped Latin American decision-makers avoid such costly errors.

Regarding the relative advantages of experimental evaluation versus statistical control, one speaker observed that the complexities of many decisions often make it impossible to rely on statistical control, since we simply do not always know all the relevant variables to measure. Another speaker stressed the need for both statistical control and experimental design, citing the continuum and the complementarity between the two. A third speaker noted Chernichovsky had discussed the differences at the *analytical* stage between experiments and statistical control, whereas the most important differences probably arise at the *data gathering* stage. Therefore, the speaker thought, it is not necessarily true that statistical control methods are always less costly than experimental control. This speaker also said his own experience indicated that existing program records are not necessarily less costly as data sources than surveys, due to the considerable costs of preparing the records for computer processing.

Several other speakers also discussed possibilities for using existing non-experimental data in evaluation studies instead of, or as complements to, new survey data. In Colombia, for example, a sample drawn from the Census files is large enough to permit analysis down to the district level of fertility, infant mortality, income, and education. Using such data for baselines, new evaluations might sometimes be limited to one-shot surveys. It was

observed that in Latin American countries, many of the following existing data categories might also be used in evaluations of health and nutrition projects: price and wage surveys, household budget surveys, agricultural production censuses and surveys, import data, children's weight charts, and research theses in anthropology and sociology by foreign professors and graduate students.

Chernichovsky remarked that he wrote his paper largely as a reaction to experiences working with nutritionists and said he thought economists were more prone than other professionals to work with existing data sources, many of which originally had no connection to economic analysis. He also noted that econometric analysis and statistical control techniques have become much less costly in recent years, due to improvements in computer technologies. And he spoke of the assistance anthropologists could render both in helping "model" family behavior and in improving data sources.

Regarding the analysis of household decision-making and the "new" economic theory of the household, one speaker emphasized that we should not view households' decisions about nutrition and health as "erroneous" simply because they do not necessarily accord with our own preconceived notions. In particular we should realize that simple educational or incentive programs will often not be enough to induce the nutrition or health behavior we want low-income households to follow, and we must be very modest in analyzing the complexities of households' decisions. Another speaker noted not only that target households often fail to respond in the ways programmers want; even worse, program resources may often go mostly to the benefit of unintended households or individuals. (But yet another speaker vigorously defended the milk distribution programs in his country, which he said had achieved an average utilization by intended families of about 60 to 70% since 1925. If there were serious distribution problems, he thought they were mainly intrafamilial).

It was observed that economists have been getting away recently from the concept of "market" or "monetary" income as a measure of the low-income family's welfare. In particular, the family's total available *time* may well be its most important resource. The family's time can be used to generate market income, but it can also be used to generate services (and commodities) like childcare that for many (especially very poor) families may be more valuable than the money income the same amount of worktime can generate (at the margin). Another speaker referred to studies that found health status better predicted by "material style of life" (i.e., various material possessions) than by measured family income.

There was also discussion to the effect that understanding household behavior, although critically important, will not be enough for proper evaluation of large social programs. We also need to do better jobs of analyzing *community* behavior. For example, it may be a social process that determines such things as who cooperates in a program, who comes to present himself for treatment, and why. Such a process may create a "selectivity bias" in household data. Therefore, we might need a large sample of different communities in order for some evaluations to achieve statistical significance. That is, communities receiving different treatments, rather than households, will be the ideal units of analysis in many cases.

There also was extensive discussion of specific outcome measures which could be considered for an evaluation program. These are summarized below.

OUTCOME MEASURES

The group split outcome measures into two levels, household and community.

Household Variables

Income. This serves as a general measure of the flow of resources available for the household to sustain its well-being. Health and nutrition may augment this flow through increases in members' productive capacities. In urban areas, where wages are the main source of income, it is fairly feasible to obtain useful data. In predominantly poor rural areas, income flow can be measured, but it requires careful attention to non-monetary dimensions. In either location, a typical questionnaire instrument might measure money income, in-kind home-produced and consumed goods and services, labor of household members, transfers from outside the household, and financial flows from assets (e.g., long land leases). Concerns were expressed about valuation and aggregation of these components. Given these problems, some expressed preference for stock measurements (see assets) rather than flows.

Assets. These may better reflect total long term resources for sustained well-being. Health and nutrition can improve the capacity to accumulate assets. Measures of major asset stocks seem feasible. Changes in such stocks may, however, be more difficult to detect. Some of the components to be measured were identified.

- a. Human capital: educational attainment and activity (time of children in school); health status; skill accumulations (training, licenses, experience).
- b. Land and productive structures: land tenure and status; land value; structures; and quality of structures. Concern was expressed that through knowledge of local conditions plus experience with specific level status measurements, an empirical estimation should be called upon for details of choices and specification of measures. Conversion to market values, where feasible, was urged.
- c. Productive assets: animals, tools.
- d. Consumer facilities or durables: water indoors or out; electricity; sewing machines, etc.

Intra-household allocation of resources. The feeling was expressed that there is a wealth of statistical experience with components of assets and their aggregation which can be called upon to enhance data collection and analysis strategies.

Since program interventions are mediated by the household, effects may appear in the form of reallocations of resources among household members and, therefore, program impacts may be transformed in unexpected ways. For some areas, measurement of intra-household allocations appeared feasible and inexpensive, for others, it appeared feasible but likely to be too expensive, and for some areas, unfeasible. Measurement and analysis could include three major areas:

- a. Individual activity: both income and asset measures call, to some degree, for recording of time allocations of household members (e.g., field time, home production time, and school activity). Data collection concerning approximate time allocations should be feasible.
- b. Diet and consumption: measurement problems could make these outcome measures too costly to be feasible for large scale use. Experience suggests that less costly methods have questionable value. Whatever measures of dietary intake or nutritional status are selected, they *should be extended to the entire household*, and not be limited just to program participants, since reallocation may shift program effects from participants to other household members.

- c. Expenditures: household budget studies must be detailed in order to capture reliably the expenditures of low income people, particularly in rural areas where non-local items are liable to be expensive. Therefore, their use must be concentrated and specialized for a given limited project evaluation, if they are used at all.

Community-Level Variables

This topic was given limited attention. Household variables should be supplemented by community-level variables: electricity; water supply and sanitation facilities; as well as differences in distance and access to community facilities. These resources maintain well-being. To the extent project interventions directly or indirectly change them, these outcomes are altered and household-level measures may not capture them adequately.

CONTROL VARIABLES

The group also spent considerable time discussing control variables and problems in their analysis. Control variables can be used: (a) to increase precision of outcome effect estimates by reducing residual variance; (b) in interactions with program variables to detect interactive effects of treatments and non-program characteristics; and (c) to reduce bias in estimating program effects in non-experimental evaluations.

Two general topics about control variables were debated at length. The first was the feasibility and advisability of using control variables to reduce selection bias in non-experimental evaluations. Some felt such evaluations may pose more risks than no evaluation at all. Others felt clear statements of the assumptions necessary for validity of conclusions under such conditions would generally reduce the risks of misuse to an acceptable level. Some urged the pressure for randomization be increased and maintained with great vigor.

The second centered on the dangers of including endogenous variables as control variables. As one example, it was noted that labor force status and health status are endogenously related, so that to use labor force status as a control in estimating program effects on health status may seriously bias estimates.

Time Lags and Sustained Effects

The group noted many outcome variables may show impacts only with long lags after intervention. In some cases, short-term changes may not be sustained. Thus, longitudinal analysis is urged where financially and logistically feasible. Changes in household outcome variables outlined above may give some indication of the degree to which any short-term changes in other (e.g., health and nutrition) outcomes may be sustained and transformed into long-term changes in welfare. Lack of changes in the household variables may help explain why short-term gains fail in certain cases to be sustained.

THE WORLD BANK

Headquarters:

1818 H Street, N.W.
Washington, D.C. 20433, U.S.A.

European Office:

66, avenue d'Iéna
75116 Paris, France

Tokyo Office:

Kokusai Building,
1-1 Marunouchi 3-chome
Chiyoda-ku, Tokyo 100, Japan



The full range of World Bank publications, both free and for sale, is described in the *World Bank Catalog of Publications*, and of the continuing research program of the World Bank, in *World Bank Research Program: Abstracts of Current Studies*. The most recent edition of each is available without charge from:

PUBLICATIONS UNIT
THE WORLD BANK
1818 H STREET, N.W.
WASHINGTON, D.C. 20433
U.S.A.