American Journal of Epidemiology
Copyright © 1991 by The Johns Hopkins University School of Hygiene and Public Health

Vol. 133, No. 2
Printed in U.S.A.

Underlying and Proximate Determinants of Child Health: The Cebu Longitudinal Health and Nutrition Study

The Cebu Study Team[1]

A proper understanding of infant health requires the integration of socioeconomic, behavioral, and biomedical models. A methodology is presented for assessing the effects of "underlying" social factors and "proximate" behavioral and biomedical factors on infant morbidity, growth, and mortality. The method is applied to data collected from over 3,000 children in Cebu, Philippines, over the first 2 years of life. Data were collected between 1983 and 1985. A central theme is that mothers recognize certain observable and nonobservable threats to the health of their infants, and that the mothers take measures to reduce the risk from such threats. It is shown that if conventional statistical techniques (which do not take such behaviors into account) are used, the estimates of the effect of the risk factors on health are incorrect. Procedures for obtaining correct estimates are described. The application of the methodology is illustrated by modeling childhood diarrhea, and by showing how maternal education induces behavioral changes, and how these changes, in turn, induce changes in the prevalence of childhood diarrhea. *Am J Epidemiol* 1991;133:185–201.

biological factors; diarrhea; epidemiologic methods; growth; health behavior; models, statistical; socioeconomic factors

The causes for high levels of childhood disease in developing countries have been the subject of numerous investigations by both social and biomedical scientists (1). The focus of the social science literature is on examining the relations between "underlying" socioeconomic variables and health outcomes, with most research focusing on mortality (e.g., references 2–6). Many of the results of these analyses are robust, with increased household income and maternal education, for example, consistently emerging as powerful determinants of health (e.g., references 7, 8). However, this literature usually gives rise to conclusions which are so sweeping (such as "where income and edu-

Received for publication November 16, 1988, and in final form February 9, 1990.

[1] The Cebu Study Team members and their affiliations are as follows: Linda S. Adair, John S. Akin, David K. Guilkey, and Barry M. Popkin (the University of North Carolina and the Carolina Population Center, Chapel Hill, NC), John Briscoe (the World Bank, Washington, DC); Robert E. Black (The Johns Hopkins University, Baltimore, MD); and Wilhelm Flieger (the Office of Population Studies of the University of San Carlos, Cebu, Philippines) John Briscoe had principal responsibility for organizing and writing this paper.

Reprint requests to John Briscoe, World Bank, 1818 H Street, N.W., Washington, DC 20433.

This paper is part of a collaborative research project involving the Office of Population Studies, University of San Carlos, Cebu, Philippines, directed by Dr. Wilhelm Flieger; the Nutrition Center of the Philippines, directed by Dr. Florentino S. Solon; and a group from the Carolina Population Center. Barry M. Popkin of the University of North Carolina is the project coordinator. Funding for parts of the project design, data collection, and computerization was provided by the National Institutes of Health (contract nos. R01-HD19983A, R01-HD23137, and R01-HD18880), the Nestles' Coordinating Center for Nutrition Research, Wyeth International, the Ford Foundation, the United States National Academy of Sciences, the Carolina Population Center, the United States Agency for International Development, and the World Bank. David Fugate, Deborah Barnes, Eilene Bisgrove, Frances Dancy, Patricia Dargent, Lionel Deang, Maria E. Fernandez, Gina Gantt, Amanda Lyerly, Christine Moe, James Vanderslice, and Emelita Wong provided invaluable assistance at many stages of the work.

cation are higher, health is better") that they provide little guidance to those concerned with formulating health programs. The mechanisms by which the socioeconomic determinants affect health remain largely an unexplored and unexplained "black box" (1).

The biomedical literature, on the other hand, focuses on the biologic precursors (such as infection and malnutrition) of morbidity and mortality. The virtue of the biomedical literature (namely its reliance on a biologic model and the focused nature of the questions answered) is its curse, too. This is so because: these narrowly focused studies have often ignored the effects of important confounding variables (with the biomedical literature on breast feeding being a good example); the ultimate consequences for mortality in populations at large tend to be neglected; and the fact that people perceive threats to their health and react to these by changing their behavior is often either not recognized or ignored because it is considered analytically intractable. The result is a literature which inevitably leads to policy conclusions favoring strictly medical interventions (1).

Drawing heavily on an analogous situation in the field of fertility research (9), Mosley and Chen (1) and Mosley (10, 11) have argued for the development of a new approach to child health research which incorporates both the social and biomedical approaches into a coherent analytic framework in which the relations between "underlying," "intermediate," and "outcome" variables are investigated. Important steps have been taken in recent years in studies in Malaysia (12–14) and Jordan (15) to conduct empirical research on mortality using this framework. The Cebu study was designed to build on these landmark studies.

This paper describes the methodology used in modeling child health in the Cebu study. Empirical results are given to illustrate the usefulness of the approach. Detailed discussions of the results and their implications have been presented in other papers (16–18).

THE DESIGN OF THE CEBU STUDY

The principal objective of the Cebu study was to correctly estimate the effects of underlying and proximate determinants of child health. Data were collected in the metropolitan area around the city of Cebu in the central Philippines. After a pilot study, a stratified, single-stage sampling procedure was used to select 17 of 158 urban and 16 of 85 rural neighborhoods in the metropolitan Cebu area. Households were surveyed to collect data on all births between May 1, 1983 and April 30, 1984. The sample consisted of 3,080 women (77 percent of whom were urban) having single live births, for whom both baseline pregnancy surveys and birth information are available. Participation rates were high. Over the course of the 2-year period, 311 of the 3,080 women (264 of 2,355 in urban areas) were lost as a result of migration, and 49 of the mothers (39 in urban areas) decided to withdraw from the study.

For each study child, questionnaires were administered in the third trimester of pregnancy, at birth, and at 2-month intervals through the first 2 years of life. Where necessary, the questionnaires were supplemented by observations (e.g., of sanitary conditions) and measurements (e.g., of weight and water quality). Information was collected on "underlying variables" (including family income and assets, education of family members and other socioeconomic variables, prices of foods and other goods and services in the community, and accessibility to health facilities), "intermediate variables" (describing households' consumption choices for health-related goods and services, such as prenatal care and infant feeding patterns, water-use practices, personal hygiene practices, use of preventive health services, maternal smoking and drinking) and "outcome variables" (including gestational age and birth weight, and growth, morbidity, and mortality at each subsequent 2-month interval). Additional details on the survey design and data are available (16–18).

MODELS FOR ASSESSING THE EFFECT OF UNDERLYING AND INTERMEDIATE VARIABLES ON CHILD HEALTH

The mechanisms whereby socioeconomic, behavioral, and biomedical factors affect health can be described in terms of two sets of equations. The first equation describes how the underlying individual, family and community variables determine health-related behaviors; the second equation describes how underlying and intermediate behavioral and biomedical variables affect health outcomes. Following standard economics terminology, these are referred to as "structural equations." For reasons which will become apparent later it is necessary to distinguish between "endogenous variables," whose values are determined by forces operating within the model, and "exogenous variables," whose values, while important to the model, are determined by forces outside the model and are not explained by the model. In the present context (as shown in figure 1 and table 1), variables such as infant feeding patterns, use of medical facilities, type of water supply and sanitation, maternal work status, and health status of the child are treated as endogenous, while variables that are not the result of health-related household decisions (such as maternal education and food prices) are considered exogenous.

The variables entering the models (where the subscript "i" refers to the particular child and the subscript "t" to the time period) are: H_{ti}, the health of the infant; Y_{ti}, endogenous variables measuring the consumption of health-related goods; Z_{ti}, exogenous community and household characteristics; μ_i, an individual-specific disturbance term that does not change through time; and ϵ_{ti}, purely random errors that vary across individuals and through time.

Structural equation 1: Determinants of behavior

The underlying family and community variables (the Z_is) are hypothesized to determine the health-related behaviors (Y_is) as follows:

$$Y_{ti} = \alpha_1 H_{Gt-1,i} + \alpha_2 H_{St-1,i} + \alpha_3 Y_{t-1,i} + \alpha_4 Z_{ti} + \mu_{Yi} + \epsilon_{Yti} \qquad \text{(expression 1)}$$

$$\text{for } i = 1, 2, \cdots, N;$$

$$t = 2, 4, \cdots 24 \text{ months}$$

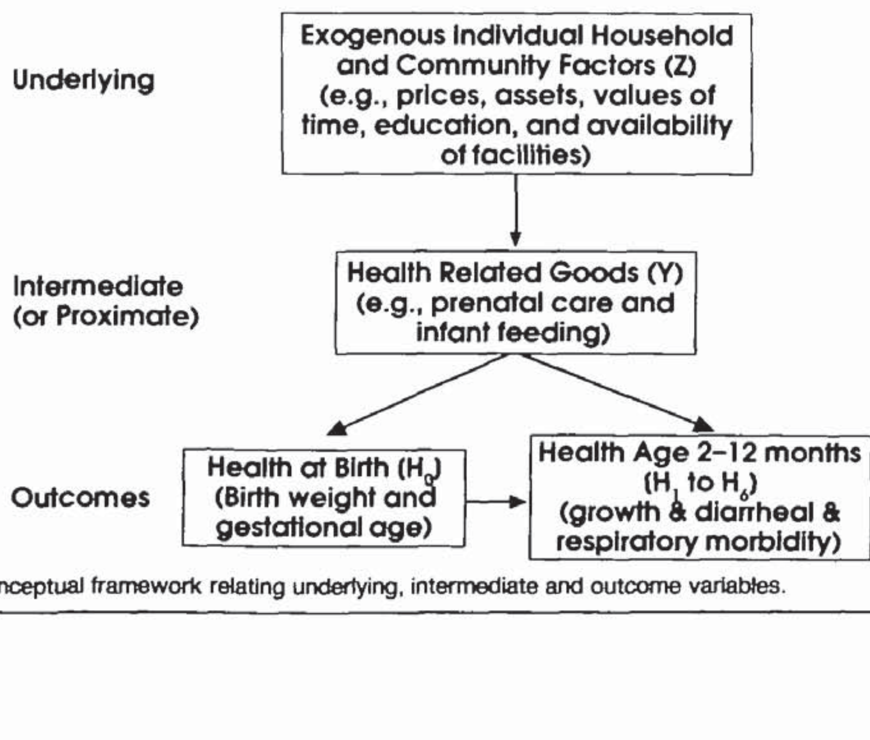

Figure 1. Conceptual framework relating underlying, intermediate and outcome variables.

TABLE 1. Endogenous variables in the diarrhea model for Cebu, Philippines, 1983–1985

I. Health of the infant
  A. At birth (H_0)
    1. Birth weight [continuous, OLS]*
    2. Gestational age [continuous, OLS]
  B. Every 2 months (H_t, $t = 2, 4, \ldots, 12$)
    1. Diarrhea [dichotomous, probit]
    2. Weight [continuous, OLS]
    3. Severe Respiratory Infection [dichotomous, probit]
II. Health-related factors affecting H_t ($t = 2, 4, \ldots, 12$) (Y_t)
  A. Exclusive breast-feeding pattern [dichotomous, probit]
  B. Any breast-feeding pattern [dichotomous, probit]
  C. Infant supplemental food nutrient intake [continuous, OLS]†
  D. Preventive medical care for infant [dichotomous, probit]
  E. Personal hygiene (soap use) [continuous, OLS]
  F. Food processing [dichotomous, probit]
  G. Immunizations (2 DPT‡; 3 DPT; measles) [dichotomous, probit]
  H. Water quality [dichotomous, probit]
  I. Excreta disposal [ordered discrete, ordered probit]

[ ] Indicates type of variable and estimation technique.
* OLS, ordinary least squares regression.
† During the period of exclusive breast feeding, a tobit estimation procedure was used.
‡ DPT, diphtheria-pertussis-tetanus vaccination.

and where the subscripts "G" and "S" refer to growth and sickness (or morbidity), respectively.

Note that the health variables measured before the current time period (the "lagged" health variables) have an effect on current health and non-health-related behaviors. That is, the model takes account of the fact that a mother might, for instance, alter her infant-feeding practices if her child failed to grow in an earlier time period. Note, too, that the model allows for the possibility that children and families have peculiarities which have important effects on how a child is treated, but which cannot be observed. For example, some mothers may know from prior pregnancies that their children are likely to be frail, and may take account of this when deciding whether to seek prenatal care.

The formal procedure for taking this unobserved heterogeneity into account is to have two (rather than the usual one) disturbance terms in the behavioral equations. The first disturbance term, μ, is specific to the individual and does not change through time. In these models, μ represents the initial endowments of the infant which cannot usually be observed by the researcher (such as "frailty"). The second error term, the ϵ, is the conventional random error term.

Structural equation 2: Determinants of health

The second structural equation describes how underlying and intermediate behavioral and biomedical variables affect health outcomes. It is assumed that the probability of being sick (e.g., of having diarrhea) in a particular period is determined in part by the nutritional status in the prior period, and that growth in a particular period is determined in part by the morbidity experience in the preceding period. In addition, it is assumed that both nutritional status and morbidity in any particular child in any particular period are also affected by the health-related and non-health-related behaviors in the prior period and by the family characteristics in the current period (the Z_ts). Accordingly, the "health structural equations" are:

$$H_{Gti} = \theta_1 H_{Gt-1,i} + \theta_2 H_{St-1,i} + \theta_3 Y_{t-1,i} + \theta_4 Z_{ti} + \mu_{Gi} + \epsilon_{Gti}$$

$$H_{Sti} = \gamma_1 H_{Gt-1,i} + \gamma_2 Y_{t-1,i} + \gamma_3 Z_{ti} + \mu_{Si} + \epsilon_{Sti} \qquad \text{(expression 2)}$$

$$H_{Mti} = \delta_1 H_{Gt-1,i} + \delta_2 H_{St-1,i} + \delta_3 Y_{t-1,i} + \delta_4 Z_{ti} + \mu_{Mi} + \epsilon_{Mti}$$

$$\text{for } i = 1, 2, \cdots, N;$$

$$t = 2, 4, \cdots 24 \text{ months}$$

and where the subscripts "G," "S," and "M" refer to growth, sickness, and mortality, respectively.

As a specific example, consider the structural equation for diarrhea. One measure of diarrhea from the available data is the dichotomous variable indicating whether the child had diarrhea in the 7-day period preceding the interview. The endogenous variables include exposure variables (water quality, personal hygiene practices, excreta disposal practices, food hygiene practices, whether exclusively breast-fed, mother's concern with preventive measures as measured by use of a well-baby clinic) and susceptibility variables (nutritional status as measured by weight for age, whether breast-fed at all, and whether vaccinated against measles), while the exogenous variables include exposure to animals in the home, season, sex, household size, community density and child's mobility.

The "reduced form" equations

The two sets of structural equations represent a complete description of how behavior and health are determined. By starting at birth and continuously substituting out for all endogenous right-hand side variables, the so-called "reduced form" equations are obtained.

$$Y_{ti} = \sum_{r=o}^{t} Z_{t-r,i}\, \beta_r^Y + \mu^*_{Yi} + \epsilon^*_{Yti}$$

$$H_{Gti} = \sum_{r=o}^{t} Z_{t-r,i}\, \beta_r^G + \mu^*_{Gi} + \epsilon^*_{Gti} \qquad \text{(expression 3)}$$

$$H_{Sti} = \sum_{r=o}^{t} Z_{t-r,i}\, \beta_r^S + \mu^*_{Si} + \epsilon^*_{Sti}$$

$$H_{Mti} = \sum_{r=o}^{t} Z_{t-r,i}\, \beta_r^M + \mu^*_{Mi} + \epsilon^*_{Mti}$$

In the reduced form equations, the endogenous variables (the Ys) and the outcome variables (the Hs) can be expressed just in terms of the exogenous variables (the Zs). These reduced form equations may be used to examine the full effect of exogenous variables (such as maternal education) on child health. They are also used in estimating the parameters of the structural equations.

PROBLEM 1: ESTIMATING PARAMETERS WHEN SOME VARIABLES ARE ENDOGENOUS

The standard approach would be to estimate the parameters of such models using standard statistical procedures (such as ordinary least squares or logistic regression). This section shows that because people recognize (some) health threats and take measures to reduce their risk from these threats, these standard procedures give the wrong answers. The correct statistical procedures are described.

Statistical procedures

For the reduced form equations (expression 3), it can reasonably be assumed that the "heterogeneity disturbance terms" (the μs) are distributed independently of the exogenous variables (the Zs) and that the error terms (the μs and the ϵs) are normally distributed. Accordingly, standard statistical techniques (such as ordinary least squares or logistic regression) can be used to obtain unbiased estimates of the parameters of the reduced form equations.

It is, however, the structural equations (expressions 1 and 2) that are of most interest to policymakers, because they permit assessment of the effects of different social, economic, and biomedical interventions on behavior and health. These equations present more complex estimation problems. For example, in expression 3 it can be seen that the value of, say, Y_{ti} depends on, and is therefore correlated with, the "heterogeneity disturbance term" (μ_{Yi}). Since the value of μ_{Yi} does not change over time, μ_{Yi} is similarly correlated with Y_{t-1}. In expression 1, therefore, a regressor ($Y_{t-1,i}$) is correlated with a disturbance term (μ_{Yi}). Similarly in expression 2 we also have a regressor ($Y_{t-1,i}$) correlated with a disturbance term (μ_{Gi}).

If the parameters of expressions 1 and 2 are estimated using ordinary least squares, inconsistent estimates will result. This is because, in explaining the dependent variable, as much credit as possible is given to the regressor and as little as possible to the error. When the regressor and error are correlated, some of the effect of the error is wrongly attributed to the regressor (19).

A standard procedure for dealing with this problem is that of "instrumental variables" (20). In estimating the parameters of the structural equation, the value of the problematic regressor (such as Y_{ti}) is not used, but replaced by an instrumental variable. The instrumental variable for Y_{ti} is chosen so that it is correlated with the regressor (Y_{ti}) and is uncorrelated with the disturbance term (μ_{Yi}). In this particular case, the "predicted values" of the Ys obtained from versions of the reduced form equations (expression 3) can be used as instrumental variables for the Ys in the structural equations (expressions 1 and 2). Using these predicted values rather than the actual values of the endogenous variables, standard estimation procedures are used to estimate the parameters of the structural equations. Although not widely known in the epidemiologic literature, such techniques are used routinely by economists (21).

As discussed in more detail elsewhere (16–18), in the Cebu study the models were specified to include all endogenous and exogenous variables, and to allow each of these to vary over time (by including a time interaction term). Only those interactions which were statistically significant were retained. The "random effects" estimation procedure (21) was used to estimate the parameters of the models. It was assumed that the μs and ϵs are normally distributed random variables and that corresponding to the observed dependent variables is a continuous latent variable (the severity of the child's diarrhea in this case). The mother reports that the child has diarrhea if the latent variable is sufficiently large (see reference 22 for a different child health model that also uses latent dependent variables). The actual calculations were done with the HOTZTRAN algorithm (23), using maximum likelihood methods in which the distribution of the disturbance term is taken into account in calculating the standard errors of the coefficients.

Does endogeneity make a difference in practice?

Two examples from the literature. The US Environmental Protection Agency has conducted a large-scale study on the effects of medical care and air pollution on mortality from respiratory disease (24). Whereas prior epidemiologic studies had implicitly assumed that people accept air pollution passively, the study recognizes that "people have an incentive to adapt to environmental conditions (by incurring the expense of seeing a doctor or moving away from a polluted city)" (25, p. 42). In analytic terms, this means that "protective factors" (such as use of medical care) are not exogenous (i.e., determined by forces outside of the model) but endogenous (i.e., determined by the level of air pollution and other factors incorporated into the model). The study (24, 25) shows, first, that the conclusion drawn from a conventional analysis was that the level of medical care had no effect on mortality from respiratory illness, but, second, that when statistical procedures taking account of endogeneity were used, medical care was shown to have a significant protective effect.

The second example deals with the effect of prenatal care on child health. A conventional analysis would treat the quantity of prenatal care as an exogenous variable and examine the relation between the level of this variable and infant health. In fact, however, many mothers seek prenatal care in part because they perceive (for reasons that are valid but which investigators cannot observe) their fetus to be particularly vulnerable. A detailed assessment of the relation between prenatal care and infant mortality in the United States (3, 5) has shown that a conventional analysis (which ignores this behavioral aspect) would conclude that mothers place their children at risk by obtaining prenatal care, but that when statistical procedures take account of this behavioral relation, use of prenatal care is shown to have a strong protective effect.

Later in this paper, the practical consequences of ignoring endogeneity are examined for the Cebu study.

PROBLEM 2: SAMPLE SELECTIVITY IN LONGITUDINAL STUDIES

As children are followed over time, there are inevitably losses to the study from refusal, out-migration, and death. Since children with certain characteristics are more likely to be lost to the sample than other children, the sample has been reduced in a way which is certainly not random. An exogenous factor which affects migration (such as father's occupation) but does not affect child health would tend to emerge from the analysis as a determinant of child health.

The procedure for correcting for this possibility consists of introducing a "correction factor" (technically known as the hazard rate or the inverse of the Mills' ratio (20, 26)) which is equal to zero for those individuals who would, without any doubt, remain in the sample throughout the period, and is relatively large for those who are likely to have been lost from the sample during the period. The statistical procedure involves, first, determining whether there is significant self-selection (if the "correction factor" is significant) and, if so, applying the necessary correction.

AN ILLUSTRATIVE RESULT: THE PATHWAYS THROUGH WHICH MATERNAL EDUCATION AFFECTS BEHAVIOR AND CHILD HEALTH

The Cebu Study Group has already published detailed results from some early analyses (16–18). For the present purposes, some empirical results illustrate how the model may be used to assess the biomedical and socioeconomic determinants of child health. The example chosen is one of major policy interest because of the consistent and strong relation (7, 8) between maternal education and child health and because of the paucity of data delineating the mechanisms by which this effect operates. The example is developed only for the urban sample, only for diarrhea, and only for the first year of life.

Underlying-proximate relations

The parameters of the behavioral structural equations (expression 1) are estimated for each of the health-related behaviors at each particular stage of the child's life. Table 2 shows the effects of maternal education on health-related behaviors during each 2-month period. Table 3 shows the simulated effects on the mean values of the health-related behaviors of increasing the education of each woman in the sample by one year.

Question 1: Is the direction of the effect sensible? Table 2 shows that as maternal education increases, there are increases in food intake, preventive health care, measles

TABLE 2. Effects of maternal education on health-related behavior during the first year of life, urban Cebu, Philippines, 1983–1985

Health-related behavior	Characteristics of dependent variable	Age of infants (months)					
		2	4	6	8	10	12
Feeding practices							
Breast feeding							
Exclusive breast feeding, no exposure to pathogens, 7 days before survey	Binary†	−0.07***	−0.07***	−0.12***	−0.02	N/A	N/A
Any breast feeding, 7 days before survey	Binary†	−0.10***	−0.11***	−0.11***	−0.10***	−0.10***	−0.10***
Food intake							
Total calories	Calories	23.02***	27.93***	23.97***	23.18***	22.96***	24.81***
Health service use							
Preventive health care	Binary‡	0.05***	0.03***	0.02*	0.04***	0.04***	0.04***
Measles immunization	Binary‡	N/A	N/A	N/A	N/A	0.07**	0.06***
Health practices, personal and environmental							
Pathogenicity of food processing	Binary§	0.0004	−0.002	−0.0001	−0.0007	−0.0008	−0.002
Poor type of excreta disposal	Binary‖	−0.09***	−0.09***	−0.09***	−0.09***	−0.09***	−0.09***
Quantity of soap per capita	Grams	1.76***	1.76***	1.76***	1.76***	1.76***	1.76***
Good quality of drinking water source	Binary¶	0.002	0.002	0.002	0.002	0.002	0.002

Entries are the coefficients of mother's formal education in years in creating instrumental variables for the dependent health-related variables. The asterisks indicate the significance level for testing whether the coefficient is zero: * $\alpha = 0.10$, ** $\alpha = 0.05$, *** $\alpha = 0.01$. N/A, not applicable.

† If the infant is breast-fed, the variable is set to 1 and 0 otherwise

‡ If the infant had a preventive health care visit or had the specified type and dosage of immunization, the variable takes on the value of 1 and 0 otherwise.

§ If food processing is severely pathogenic, the variable is set to 1 and 0 otherwise.

‖ If the household's excreta disposal is poor, the variable is set to 1 and 0 otherwise.

¶ If the source of drinking water is good, the dependent variable takes on the value of 1 and 0 otherwise.

immunization, adequacy of excreta disposal practices, quantity of soap used per capita, and quality of drinking water; and decreases in breast feeding (both exclusive and any) and food contamination risk.

Question 2: Is the effect statistically significant? From table 2 it can also be seen that, for most health-related practices, the effects of maternal education are highly statistically significant. The two exceptions are unhygienic food preparation practices and quality of drinking water. In both cases, the lack of significance is almost certainly because, with the measures employed in these early analyses, there is little variation in these variables in the urban sample.

Question 3: Are the findings of practical significance? From table 3, it can be seen that a one-year increase in the education of each mother would have substantial effects on most of the health-related behaviors. For example, for a 6-month-old child, a one-year increase in maternal education implies a 36 percent reduction in the probability of exclusive breast feeding, a 5 percent reduction in the probability of any breast feeding, a 7 percent increase in caloric intake, a 4 percent increase in the use of preventive health care, a 9 percent reduction in the probability of inadequate excreta disposal practices, and a 2 percent increase in per capita soap use.

Proximate-outcome relations

The second set of structural equations (expression 2) describes the relations between the proximate (behavioral and biomedical) variables and health outcomes. Table 4 presents the proximate-outcome structural equation for diarrhea for the longitudinal model estimated for the full first year of life for the urban population. Table 5 shows the responsiveness of diarrhea to changes in the proximate variables (as measured by the "elasticity," that is, the percent change in diarrhea resulting from a percent change in the proximate variable).

Substantive Question 1: Are the estimates sensible and statistically significant? From table 4, it can be seen that diarrhea is statistically significantly lower for: faster growing infants (with the effect greatest in small infants); infants who are breast-fed (with the protective effect greatest at young ages); infants who are exclusively breast-fed; infants who consume more food; infants whose families use better quality water; infants whose families follow hygienic food preparation practices; and infants whose families have better excreta disposal practices (with the effect greater in the early months of this first year of life). Diarrhea is statistically significantly higher for: male infants in the latter months; crawling infants when there are animals in the house; and children in more densely settled communities.

Substantive Question 2: Are the findings of practical significance? From table 5, it is evident that the level of diarrhea is highly responsive to breast-feeding practices (especially in the early months of life) and to excreta disposal and water supply practices (throughout the first year of life), moderately responsive to caloric intake, especially later in the first year of life, and largely unaffected by preventive health care.

Methodological Question 1: What are the consequences of sample selectivity? Statistical analysis showed that the hazard rate (or inverse of the Mills' ratio) was small, and not significantly different from zero. It was therefore concluded that sample selectivity was not significant (that is, that nonresponse could be viewed as a random event in the sample) and that the Mills' ratios could be excluded in the final specifications of the instrumental variables.

Methodological Question 2: What are the consequences of ignoring endogeneity? The importance of taking account of endogeneity when modeling child health was tested in two ways using the Cebu data set. The first test is a formal statistical test—a chi-square version of the Hausman test (21, 27)—which indicates whether endogeneity was actually present. The critical value for a 1 percent test is 29, while the test statistic was 98: the null hypothesis of no endogeneity is strongly rejected. The results of a second, more intuitive, test (comparing results from two estimation procedures, one taking account

TABLE 3. Simulated effects of one-year increase in maternal education on health-related behavior and diarrhea incidence, at every 2 months during the first year of life, urban Cebu, Philippines, 1983–1985*

	Age of infants (months)											
	2		4		6		8		10		12	
	Mean	% Change	Mean	% Change	Mean	% Change	Mean	% Change	Mean	% Change	Mean	% Change
Feeding practices												
Breast feeding												
Exclusive breast feeding, no exposure to pathogens, 7 days before survey	0.16	−10	0.06	−12	0.0002	−36	N/A		N/A		N/A	
Any breast feeding, 7 days before survey	0.85	−3	0.79	−4	0.74	−5	0.70	−5	0.64	−6	0.57	−7
Food intake												
Total calories	121.70	10	192.60	9	307.70	7	372.00	6	439.40	5	507.50	5
Health service use												
Preventive health care	0.12	9	0.18	5	0.18	4	0.13	7	0.11	8	0.08	7
Measles immunization	N/A		N/A		N/A		N/A		0.01	19	0.01	16
Health practices, personal and environmental												
Pathogenicity of food processing†	0.04	1	0.03	−8	0.03	0	0.03	−2	0.03	−3	0.04	−5
Poor type of excreta disposal†	0.52	−9	0.52	−9	0.52	−9	0.52	−9	0.52	−9	0.52	−9
Quantity of soap per capita†	81.12	2	81.12	2	81.12	2	81.12	2	81.12	2	81.12	2
Good quality of drinking source†	0.99	0	0.99	0	0.99	0	0.99	0	0.99	0	0.99	0

* Simulation is done using the instrumental variable equation. The % change is obtained with this formula: % change = $(M_2 - M_1)/M_1 \times 100$, where M_1 is the simulated mean of the health-related behavior of interest obtained by multiplying the coefficients of the exogenous variables by their corresponding sample means. M_1 is entered as mean in the table. M_2 is the simulated mean when all variables except woman's education are at their sample means and mean maternal education is increased by one year. N/A, not applicable because no infants had this feeding or immunization during this period.

† Current version of these variables are not time-varying yet.

TABLE 4. Longitudinal analysis: Structural equation for diarrhea incidence in week preceding survey, urban Cebu, Philippines, 1983–1985

Explanatory variables	Coefficient	(t statistic)
A. ENDOGENOUS		
Susceptibility		
Lagged weight velocity (g/day)	−0.01	(−2.02**)
Lagged weight velocity interacted with weight (g × g/day)	3.30×10^{-6}	(2.90***)
Gestational age (weeks)	0.01	(3.12***)
Gestational age interacted with age (weeks × days)	9.60×10^{-6}	(1.68*)
Susceptibility/exposure		
Feeding practices		
Any breast feeding 7 days before survey (prob)†	−0.68	(−2.51**)
Any breast feeding interacted with age (prob × days)	1.60×10^{-3}	(1.76*)
Exclusive breast feeding with no exposure to pathogens, 7 days before survey (prob)	−1.53	(−5.91***)
Total calories (cal)	-4.40×10^{-4}	(−1.73*)
Exposure		
Health service use		
Preventive health care (prob)	−0.24	(−1.27)
Health practices, personal and environmental		
Good quality water source (prob)	−0 32	(−3.35***)
Soap purchased/capita/week (g)	-3.70×10^{-5}	(−0.06)
Pathogenic food processing (prob)	0.91	(1.85*)
Poor excreta disposal (prob)	0.92	(4.97***)
Poor excreta disposal interacted with age (prob × days)	-1.80×10^{-3}	(−2.45**)
B. EXOGENOUS		
Susceptibility		
Child's age (days)	7.10×10^{-4}	(1.00)
Child's sex (0–1)	−0.02	(−0.24)
Child's sex interacted with age (0–1 × days)	5.60×10^{-4}	(2.09**)
Exposure		
Animals in the house (0–1)	-6.90×10^{-3}	(−0.22)
Animals under the house (0–1)	−0.02	(−0.53)
Baby crawling interacted with animals in the house (0–1)	0.08	(1.93*)
Crowding		
No. of preschoolers (0–6)	−0.03	(−2.20**)
No. of persons/room (0–9.5)	9.90×10^{-3}	(1.05)
Community density (persons/km²)	6.50×10^{-6}	(7.09***)
Cumulative rainfall in last 2 weeks before survey (mm)	2.05×10^{-4}	(0.45)
Cumulative rainfall interacted with age (mm × days)	1.20×10^{-6}	(0.69)
C. OTHERS		
Constant	−6.05	(−3.63***)
Rho	0.12	(7.22***)

Note: Sample size for this analysis is 11,807. The significance levels for testing whether the coefficient is zero are indicated by: * $\alpha = 0.10$, ** $\alpha = 0.05$, *** $\alpha = 0.01$.

† Prob, the predicted probability of the explanatory variable

of endogeneity and one ignoring it) are presented in table 6 and summarized in figure 2.

"Column 1" of table 6 presents the results of the analysis which ignores endogeneity. For this analysis, observations are needed on all of the variables (both those which are considered exogenous and endogenous in the analysis using instrumental variables). A total of 6,674 observations are available. "Column 2" of table 6 presents the results of the instrumental variable analysis, for this same sample. A comparison of columns 1 and 2 shows that the analysis which does not account for endogeneity is reasonably specific. In only one case—total calories—does

TABLE 5. Percent change in diarrhea for a 1% increase in explanatory variables during the first year of life, urban Cebu, Philippines, 1983–1985*

	Age of infants (months)					
	2	4	6	8	10	12
Feeding practices						
Exclusive breast feeding, no exposure to pathogens	−0.59	−0.33	−0.05	N/A	N/A	N/A
Any breast feeding	−1.08	−0.91	−0.76	−0.70	−0.66	−0.60
Total calories	−0.12	−0.15	−0.22	−0.24	−0.29	−0.34
Health service use						
Preventive health care	−0.07	−0.09	−0.08	−0.06	−0.05	−0.04
Measles immunization	N/A	N/A	N/A	N/A	−0.02	−0.03
Health practices, personal and environmental						
Pathogenicity of food processing	0.08	0.05	0.04	0.05	0.04	0.06
Poor type of excreta disposal	0.90	0.71	0.54	0.43	0.34	0.24
Quantity of soap per capita	−0.01	−0.01	0	0	0	0
Good quality of drinking water source	−0.55	−0.49	−0.44	−0.43	−0.44	−0.43

* Entries are computed by the following formula: % change = $(D_2 - D_1)/D_1 \times 100$, where D_1 is the simulated mean diarrhea when the coefficients of the explanatory variables in the diarrhea structural equation are multiplied by their corresponding sample means. D_2 is the simulated mean when the value of the variable of interest is increased by 1.01 of its sample mean and the rest of the explanatory variables are at their means. N/A, not available because no infant had this type of feeding or immunization during this period.

TABLE 6. The effect of ignoring endogeneity: Structural equation for diarrhea incidence in week preceding survey, urban Cebu, Philippines, 1983–1985

Explanatory variables	Column 1		Column 2	
	Estimates when endogeneity is ignored	(t-statistic)	Instrumental variable estimates	(t-statistic)
A. ENDOGENOUS				
Susceptibility				
Lagged weight velocity (g/day)	0.08	(1.39)	−0.01	(−1.20)
Lagged weight velocity interacted with weight (g × g/day)	-2.40×10^{-6}	(−0.24)	4.10×10^{-6}	(2.13**)
Gestational age (weeks)	0.18	(0.74)	0.09	(1.48)
Gestational age interacted with age (weeks × days)	-3.20×10^{-6}	(−0.29)	-1.50×10^{-6}	(−1.41)
Susceptibility/exposure				
Feeding practices				
Any breast feeding 7 days before survey (prob)†	−0.68	(−5.58***)	−1.06	(−2.29**)
Any breast feeding interacted with age (prob × days)	2.00×10^{-3}	(3.93***)	4.30×10^{-3}	(2.20**)
Exclusive breast feeding with no exposure to pathogens, 7 days before survey (prob)	−0.23	(−1.46)	−1.35	(−3.64***)
Total calories (cal)	-2.30×10^{-4}	(−2.94***)	-3.10×10^{-4}	(−0.84)

Exposure				
Health service use				
Preventive health care (prob)	0.02×10^{-3}	(−0.31)	−0.06	(−2.30**)
Health practices, personal & environmental				
Good quality water source (prob)	−0.17	(−1.87*)	−0.22	(−1.70*)
Soap purchased/capita/week (g)	-1.20×10^{-6}	(−0.37)	3.00×10^{-6}	(0.35)
Pathogenic food processing (prob)	0.15	(1.30)	0.87	(1.27)
Poor excreta disposal (prob)	0.09	(1.04)	1.11	(3.68***)
Poor excreta disposal interacted with age (prob × days)	2.70×10^{-3}	(0.07)	-3.60×10^{-3}	(−2.71***)
B. EXOGENOUS				
Susceptibility				
Child's age (days)	2.80×10^{-4}	(0.65)	1.30×10^{-4}	(0.98)
Child's sex (0–1)	−0.05	(−0.47)	−0.07	(−0.70)
Child's sex interacted with age (0–1 × days)	6.50×10^{-4}	(1.40)	7.10×10^{-4}	(1.55)
Exposure				
Animals in the house (0–1)	−0.02	(−0.39)	-6.60×10^{-3}	(−0.12)
Animals under the house (0–1)	−0.01	(0.05)	−0.03	(−0.68)
Baby crawling interacted with animals in the house (0–1)	0.02	(1.04)	−0.01	(0.13)
Crowding				
No. of preschoolers (0–6)	0.03	(0.14)	−0.03	(−1.49)
No. of persons/room (0–9.5)	0.02	(1.15)	0.01	(0.73)
Community density (persons/km²)	5.80×10^{-6}	(5.13***)	7.60×10^{-6}	(6.26***)
Cumulative rainfall in last two weeks before survey (mm)	−0.07	(−1.20)	-4.47×10^{-4}	(−0.68)
Cumulative rainfall interacted with age (mm × days)	-4.50×10^{-6}	(−1.59)	3.30×10^{-6}	(1.17)
C. OTHERS				
Constant	−1.87	(−1.85**)	−4.51	(−1.88**)
Rho	0.16	(6.27***)	0.15	(5.86***)

Note: Sample size for this analysis is 6,674. The significance levels for testing whether the coefficient is zero are indicated by. * $\alpha = 0.10$, ** $\alpha = 0.05$, *** $\alpha = 0.01$.
† Prob is the predicted probability of the explanatory variable.

		"Standard" Analysis (Ignoring Endogeneity)		
		Sign positive and significant	*Not significant*	*Sign negative and significant*
"Correct" Analysis (Accounting for Endogeneity)	*Sign positive and significant*	breastfeeding x age community density	excreta disposal weight velocity x weight	
	Not significant		weight velocity food processing soap animals in house animals under house gestational age child's age persons / room # preschoolers child's sex rainfall baby crawling x animals gestational age x age rainfall x age	total calories
	Sign negative and significant		exclusive breastfeeding preventive services excreta disposal x age	any breastfeeding water quality

Legend:

Inference from "standard" analysis would be:

correct

moderately midleading

seriously misleading

Figure 2. Inferences from the "correct" and "standard" analyses, Cebu, Philippines, 1983–1985.

the simpler model (column 1) suggest a statistically significant relation which is not significant in the "correct" model (column 2). More serious, however, are those behaviors—improved excreta disposal, preventive health care and exclusive breast feeding—which would appear to have no effect if endogeneity is ignored (column 1) but which, in fact, have strong protective effects (as shown in the "correct" analysis in column 2). Furthermore, where the analysis ignoring endogeneity gave statistically significant estimates of the correct sign (any breast feeding, water quality, community density, and breast feeding × age), the parameter values were substantially biased toward the null. In short, if endogeneity is ignored, incorrect conclusions will be drawn on the determinants of health.

Tracing the paths by which education affects health

Interesting and important as the above results are, the major potential contribution of the Cebu Project is the integration of these two levels of analysis into a single, integrated behavioral-cum-biomedical description of child health. This integration can be illustrated by tracing through the pathways by which maternal education affects health-related behavior, and how such behavior, in turn, affects health.

Before tracing this path, the aggregate effect on diarrhea of a one-year increase in maternal education can be calculated using the reduced form (expression 3). The net effect of a one-year increase in maternal education would be to reduce the incidence

TABLE 7. Percent change in diarrhea due to behavioral change induced by a one-year increase in maternal education at every 2 months during the first year of life, urban Cebu, Philippines, 1983–1985*

	Age of infants in months					
	2	4	6	8	10	12
Feeding practices						
Exclusive breast feeding, no exposure to pathogens	4.50	1.73	0.02	N/A	N/A	N/A
Any breast feeding	2.53	2.55	2.03	1.43	0.94	0.46
Total calories	−1.00	−1.18	−1.23	−1.35	−1.30	−1.34
Health service use						
Preventive health care	−0.50	−0.35	−0.21	−0.29	−0.27	−0.16
Measles immunization	N/A	N/A	N/A	N/A	0	0
Health practices, personal and environmental						
Pathogenicity of food processing	0.06	−0.35	−0.01	−0.08	−0.09	−0.23
Poor type of excreta disposal	−6.74	−4.97	−3.76	−2.96	−2.26	−1.54
Quantity of soap per capita	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01
Good quality of drinking water source	0	0	0	0	0	0

* Entries are computed with the following formula: % change = $(D_2 - D_1) \times 100/D_1$, where D_1 is the simulated mean diarrhea when coefficients of the explanatory variables in the diarrhea structural equation are multiplied by their corresponding sample means. D_2 is the simulated mean diarrhea where the mean of instrumental variable of interest is changed by increasing mean maternal education by one year and the rest of the explanatory variables are at their sample means. N/A, not applicable because no infant had this type of feeding or immunization during this period.

of diarrhea episodes by about 5 percent in each time period. (Over the first year of life, the mean incidence of diarrhea in a 7-day period increased from under 4 percent in the first 2 months to over 15 percent in the final 2 months.)

By combining the simulated effects of education on health (table 3) with the simulated effect of behavioral changes on diarrhea (table 5), the education-behavior-diarrhea pathway can be traced. Table 7 shows that the three major pathways through which maternal education affects health in this population are: a large reduction in diarrhea (about 4 percent) because of improved excreta disposal practices, with the effect being particularly strong in the early months of life; a substantial reduction in diarrhea because of the increase in calories given to the child, with the effect greater toward the end of the first year of life; and a substantial, offsetting, increase in diarrhea because of a reduction in the number of mothers who breast-fed, with reduced exclusive breast feeding most important in the early months, and reduced breast feeding most deleterious early but remaining serious throughout the first year of life.

This information is presented graphically (for 6-month-old urban children) in figure 3. From figure 3, it is evident that some pathways are not important in this population either: because maternal education has little effect on behavior (as is the case of water supply for this urban population); or because the prevalence of the particular behavior is low (only 9 percent of mothers are exclusively breast-feeding their children at this age, for instance); or because changes in the particular behavior have little effect on health in this period (such as changes in the use of soap).

SUMMARY AND CONCLUSIONS

This paper shows that an integrated socioeconomic-biomedical model of child health can be specified and the parameters estimated. The results show that if endogeneity is ignored, incorrect conclusions are

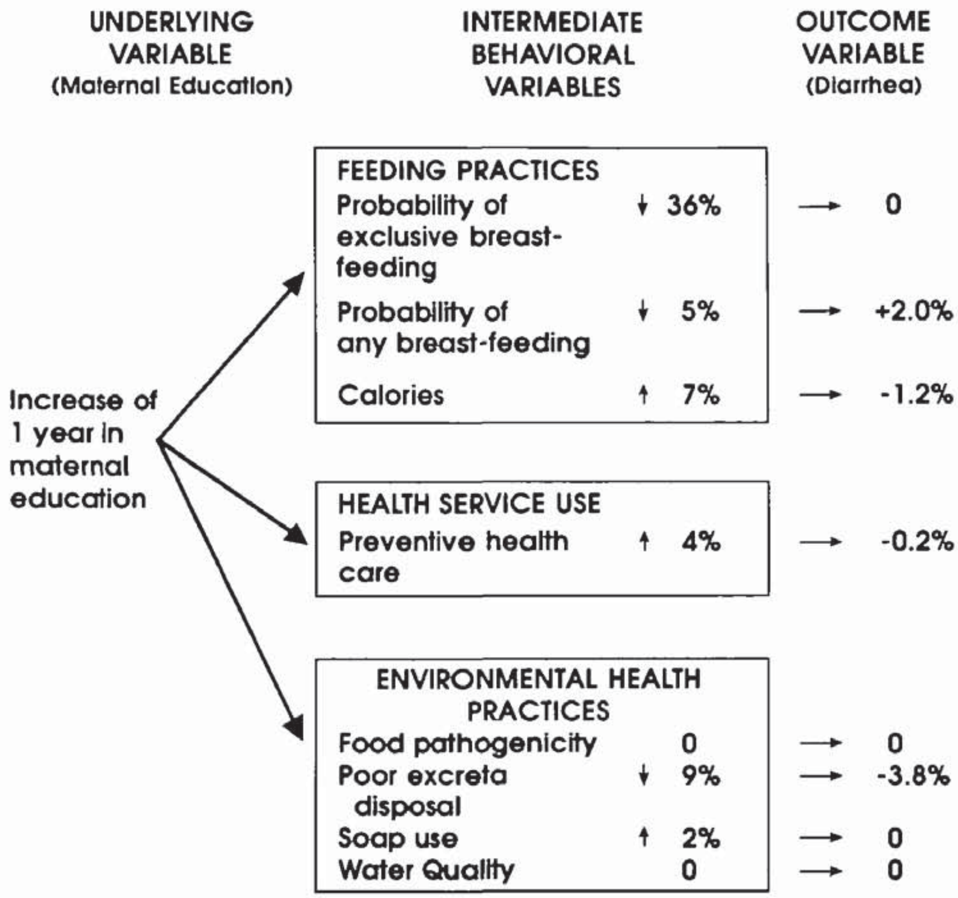

Figure 3. The pathways through which maternal education affects health at age 6 months, urban Cebu, Philippines, 1983–1985.

drawn concerning the effects of several determinants of child health. The analytic approach permits unique and readily understandable disaggregation of the effects of underlying and intermediate determinants of child health.

REFERENCES

1. Mosley WH, Chen LC. An analytic framework for the study of child survival in developing countries. Population and Development Review 1984;10:25–48.
2. Preston SH. Causes and consequences of mortality declines in less developed countries during the twentieth century. In: Easterlin RA, ed. Population and economic change in developing countries. Chicago: University of Chicago Press, 1980.
3. Schultz TP. Studying the impact of household economic and community variables on child mortality. Population and Development Review 1984;10:215–36.
4. Trussell J, Mencken J. Estimation of levels, trends and covariates of infant and child mortality in statistically poor countries. Population and Development Review 1984;10:325–46.
5. Rosensweig MR, Schultz TP. Estimating a household production function: heterogeneity and the demand for health inputs, and their effects on birthweight. Journal of Political Economy 1983;92:723–46.
6. Behm H. Socioeconomic determinants of mortality in Latin America. Proceedings of the meeting on socioeconomic determinants and consequences of mortality, el Colegio de Mexico, Mexico City, June 19–25, 1979. New York: World Health Organization, 1980.
7. Caldwell JC, McDonald P. Influence of maternal education on infant and child mortality. Proceedings of International Population Conference. Liège, Belgium: International Union for the Scientific Study of Population, 1981.
8. Cochrane SH, O'Hara DJ, Leslie J. The effects of education on health. World Bank Staff Working Paper no. 405, Washington, DC: World Bank, 1980.
9. Bongaarts J. A framework for analyzing the proximate determinants of fertility. Population and Development Review 1978;4:105–32.
10. Mosley WH. Biological contamination of the environment by man. In: Biological and social aspects

of mortality and length of life. Preston SH, ed. Liège, Belgium: International Union for Scientific Study of Population, 1980:36–67.
11. Mosley WH. The demographic impact of child survival programs. International Symposium, Center for Public Health Research, Mexico City, July 1986.
12. Davanzo J. A household survey of child mortality determinants in Malaysia. Population and Development Review 1984;10:307–24.
13. Davanzo J, Butz WP, Habicht JP. How biological and behavioral influences on mortality in Malaysia vary during the first year of life. Population Studies 1983;37:381–402.
14. Butz WP, Davanzo J, Habicht JP. Biological and behavioral influences on the mortality of Malaysian infants. Santa Monica, CA: Rand Corporation, 1982.
15. Tekce B, Shorter F. Determinants of child mortality: a study of squatter settlements in Jordan. Population and Development Review 1984;10:257–80.
16. Adair LS, Popkin B. Birth weight, maturity and proportionality in Filipino infants. Hum Biol 1988; 60:319–39.
17. Guilkey DK, Popkin BM, Akin JS, et al. Prenatal care and pregnancy outcomes in the Philippines. Journal of Development Economics 1989;30: 241–72.
18. Wong E, Popkin B, Akin JS, et al. Accessibility, quality of care and prenatal care use in the Philippines. Soc Sci Med 1987;24:927–44.
19. Wonnacott TH, Wonnacott RJ. Introductory statistics for business and economics. New York: John Wiley and Sons, 1972.
20. Judge GG, Griffiths WE, Hill RC, et al. The theory and practice of econometrics. New York: John Wiley and Sons, 1980.
21. Maddala GS. Introduction to econometrics. New York: Macmillan, 1988.
22. Behrman JR, Wolfe BL. How does mothers' schooling affect family health, nutrition, medical care usage, and household sanitation? Journal of Econometrics 1987;36:185–204.
23. Avery R, Hotz J. Hotztran users' manual. Old Greenwich, CT: CERA Economic Consultants, 1985.
24. Crocker TD, Schulze W, Ben-David S, et al. Experiments with economics of air pollution epidemiology: volume 1 of methods development for assessing air pollution control benefits. Washington DC: United States Environmental Protection Agency, 1979. (Report EPA-600/5-79-00 1a).
25. Kneese AV. Measuring the benefits of clean air and water. Washington, DC: Resources for the Future, 1986.
26. Heckman JJ. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Annals of Economic and Social Measurement 1976;5:475–92.
27. Hausman JA. Specification tests in econometrics. Econometrica 1978;46:1251–71.