5 MODEL, METHODS AND DATA

In this chapter we present the conceptual framework and theoretical model we use for our analysis. The empirical strategy following from this model is also presented. We construct variables for estimation using NFHS-3 data. The data, variables and definitions are also presented in this chapter.

5.1 INTRODUCTION

In this study, we estimate the effect of ICDS program on nutritional status of children. Parents decide how much of their time and resources to spend on procuring child health. This chapter lays out the theoretical model on which we base our analysis in section 5.2. Section 5.2 spells out the empirical strategy for estimation. The data, variables and definitions are discussed in section 5.4.

5.2 CONCEPTUAL FRAMEWORK

The theoretical approach we follow in this analysis is based on Becker's microeconomic models of household production (Becker 1965). Households allocate resources for production of goods and services that are either sold in the market or consumed at home. This work has been extensively used in many studies to estimate health or anthropometric production functions relating child height or child weight (Grossman 1972). Behrman and Deolalikar (1988), Strauss and Thomas (1995) and Currie (2003) provide excellent reviews of later applications of this model to child health status. We use the model as laid out by Strauss and Thomas (1995).

5.2.1 Theoretical Model

Assuming a simple one-period, unitary model of decision-making where a household maximizes utility

$$U_j = U(X_{jt}, L_{jt}, R_j, H_{ijt}; W_{jt}) \quad (1)$$

Utility is derived from the consumption of a set public and private goods (X_j), the consumption of leisure by all household members (L_j), child health (H_{ij}) and all other household factors (R_j) defined over household taste and preferences (W_j).

The preference function (1) is maximized subject to a set of constraints, given assets and prices. The two main constraints that the household faces are:

i. A health production function: The health of the ith child is produced by a number of child, parental and time inputs. The production of health is a dynamic process and present state of health reflects current and previous periods' inputs. The health production function of an ith child in the jth household in period t can be written as:

$$H_{ijt}= h(N_{jt}, N_{jt-1},\ldots N_{j0}, C_{jt}, C_{jt-1},\ldots C_{j0}, M_{jt},M_{jt-1},\ldots.,M_{j0}; K_{ijt}, G_{ijt}) \quad (2)$$

where health depends on N_{jt}, the inputs overtime such as nutrient intakes, immunizations etc,; on C_{jt}, the exogenous characteristics of child and household health; on M_{jt}, the time inputs (where total time available is allocated between work, leisure and health related activities; on K_{ijt}, the child related factors such as age and gender and G_{ijt}, the household related factors such as characteristics of parents, household and community factors. The last two inputs differ from the other variables in that they are not the choice variables in the period in which health decisions are being undertaken. Strauss and Thomas (1995) state that if one makes an assumption that the current state of health is a function of past health inputs, then equation (2) can be written as:

$$H_{ijt}= h(N_{jt}, C_{jt}, M_{jt}, K_{ijt}, G_{ijt}) \quad (3)$$

ii. Full-income constraint are the time and income constraints that can be written as:

$$p_t^x.X_{jt}+p_t^N N_{jt}=w_j L_{jt}+Y_{jt}+A_{jt} \quad (4)$$

where $L_{jt}= T_{jt}-S_{jt}-M_{jt}$ denotes the time-constraint and p_t^x refers to prices of private and public goods the household consumes, p_t refers to price of health related inputs, $w_j L_{jt}$ is the wage income, Y_{jt} is non-wage income, A_{jt} refer to the assets of households, T_{jt} refers to total time and S_{jt} refers to leisure time. It should be noted that L_{jt} incorporates the time endowment constraint of the household and is the time spent on earning wages (total time is spent on labor, leisure and health-related activities).

Under the assumption that the utility function is increasing and quasi-concave and the underlying functions have desirable properties so that an internal maximum can be obtained, the utility function can be maximized subject to the constraints to solve for

the optimal amount of health inputs (N_{jt}^*), health-related time inputs (M_{jt}^*), optimal amount of household consumption (X_{jt}^*) and leisure time (S_{jt}^*) to obtain reduced-form demand functions. In general, the reduced-form demand relations give some behavioral outcome in the current period as dependent on all predetermined (from the point of view of the entity making the decisions) prices and resources and on the parameters in the underlying production functions and preferences. Thus, these reduced-form demand functions can be written as

$N_{jt}^* = n(H_{ijt}, p_t^x, p_t^N, w_t, T_{jt}, Y_{jt}, A_{jt}, K_{jt}, G_{jt})$
$M_{jt}^* = m(H_{ijt}, p_t^x, p_t^N, w_t, T_{jt}, Y_{jt}, A_{jt}, K_{jt}, G_{jt})$
$X_{jt}^* = x(H_{ijt}, p_t^x, p_t^N, w_t, T_{jt}, Y_{jt}, A_{jt}, K_{jt}, G_{jt})$
$S_{jt}^* = s(H_{ijt}, p_t^x, p_t^N, w_t, T_{jt}, Y_{jt}, A_{jt}, K_{jt}, G_{jt})$

It is a characteristic of reduced-form demand models that all the exogenous prices enter into the determination of each of the endogenous variables, implying that health depends on the price of all goods and not just health-related inputs. Also, the predetermined variables enter reduced-form demand functions implying that assets, endowments of households, parental and community factors affect child health and nutritional status.

5.2.2 Empirical Model

The reduced-form demand equation that is derived is a function of individual, household and community characteristics, income and prices and program variable. We are interested in measuring the program impact on health outcomes of children and also other determinants of children's health.

For estimation, we assume that the reduced-form demand function is linear. For a particular child, the health function can be written as:

$$H_{ijt} = b_0 + b_1 Z_1 + b_2 Z_2 + b_3 ICDS + u \qquad (5)$$

where ICDS refers to the program (treatment) area, one of the possible household resources, that has been singled out because of its central interest for our analysis. We wish to obtain an estimate of the impact of ICDS on child anthropometric measures of health, i.e, a good estimate of the parameter b_3. ICDS refers to the presence of the program. Z_1 refers to all the child-related factors and Z_2 refers to a set of the parental

and household factors. u is the disturbance term. The parameters that are to be estimated are the vector of coefficients given by b. The basic estimation problem is that there are likely to be many unobserved variables that affect child nutritional status within this framework and that may be correlated with whether a particular child in a particular household is in the ICDS program area. For example, a household with access to ICDS may be more likely to take advantage of the program if the parents have greater concern about their children's welfare and future prospects, if the parents are more educated, if they belong to a particular caste or religion or if the local environment is relatively unhealthy. As a result, if there is no control for such factors, the estimated b_3 will be contaminated by omitted variable bias and may differ substantially and possibly even in sign from the true value. The estimation strategy used in this study controls for such factors. Squares terms of regressors as well as interaction terms between the covariates are introduced in the empirical specifications to account and control for any nonlinearity.

The empirical specifications involve estimations of five variants of the model:

In the base model, we include only the program variable. In the second model, the program effect controlling for only the individual characteristics is examined. We include maternal factors in the third model. The fourth model controls for all paternal and household characteristics. The full model includes all the controls including the interaction terms. We estimate these models for different age groups, gender, states and varying program definitions.

Propensity Score Matching

To tests our results further, we also use the propensity score matching (PSM) method to estimate the average treatment effects on the treated. This method is explicitly based on the assumption that, given the observable covariates, participation in the program is independent of the outcomes. That is, the observable variables capture any systematic differences between those in the program areas and those not in the program areas. Other than those differences, there would be no systematic difference in the outcome if the program had not been implemented. Out in the perspective of our analysis, we are interested in estimating the effect of ICDS program on child health outcome. If there was no possibility of bias, it would be possible to estimate the average treatment effect (ATT) i.e. the effect of the program simply by subtracting the

average health (WAZ or HAZ) outcome of children in the non-program area from average of those in the non-program area.

$$ATT = E[Y_{i1}|D=1] - E[Y_{i0}|D=1]$$

where D= dummy for program area (if area is in ICDS or not/has a AWC or not)

$Y_{i1}|D=1$ is the outcome (of child health) if there is ICDS in the area

$Y_{i0}|D=1$ is the outcome (of same child health) if there is no ICDS in the area

The above equation is analogous to an experimental study where ($Y_{i1}|D=1$) represents the group of subjects randomly assigned to the treatment (ICDS program) and ($Y_{i0}|D=1$) represents the group of subjects randomly assigned to control (no ICDS program). However, because being in the ICDS area is not random; it is not possible to determine what effect on child health would have been in an area, had it not been under the ICDS program. Moreover, ICDS program is not designed to be an experimental program to test the effect by randomly assigning areas to program or maintain them as control. Given that placement of ICDS area could be based on some observable data (based on initial guidelines) and the fact that use of these services is a choice to be made by individuals (implying a process of self-selection), there could be biased estimates and these issues need to be addresses. PSM is a technique to do the same in non-experimental studies (as also mentioned in Chapter 4).

Matching estimators assign each member of the treatment group to a member(s) in the control group who exhibits similar observable characteristics. Conditional on those characteristics, potential experimental outcomes are assumed independent of treatment or control group membership. Only those members of the treatment group and the control group for whom a suitable match can be found are included in the analysis The larger the number of the characteristics that determine the common support, the more complex and computationally intensive is the resulting matching process. Fortunately, Rosenbaum and Rubin (1983) prove that matching can be obtained by pairing treatment and control subjects according to their propensity scores obtained from a parametric model that estimates the conditional probability of belonging to an experimental treatment group.

In our analysis, we use Leuven and Sianesi's (2003) method of propensity score matching tool for STATA (PSMATCH2). The rationale for PSM is to recreate the above equation with a mock control group to simulate a randomised experiment. If we

assume that conditional on attributes X_i the outcomes are independent of the program participation then it is also possible to observe the average treatment effect with the following:

$$ATT= E [Y_{i1}|D=1, X_i]- E [Y_{i0}|D=0, X_i]$$

$Y_{i1}|D=1, X_i$ is the average outcome (health) in the ICDS areas under conditions X_i

$Y_{i1}|D=0, X_i$ is the average outcome (health) in the non-ICDS areas under conditions X_i

Following the above discussion on the intuition for propensity score matching, there are three basic steps in Leuven and Sianesi's (2003) PSMATCH2 tool. First, PSMATCH2 uses a logistic regression to predict the propensity to be in the program area. Second, PSMATCH2 matches data based on their propensity scores. Third, PSMATCH2 determines the treatment effect by averaging the difference in health outcome between those in the program (treated) and those not in the program (untreated). The covariates we use for propensity score matching consist of the child, household and community variables.

5.3 DATA AND MEASURES

5.3.1 Data

The basic data source for analysis in this thesis is the third round of National Family Health Survey (NFHS-3). The National Family Health Surveys (NFHS), initiated in the early 1990s, has emerged as a nationally important source of data on population, health, and nutrition for India and its states. These surveys are conducted by International Institute of Population Sciences under the auspices of Ministry of Health and Family Welfare (MoHWF). ORC Macro under the Demographic and Health Surveys (DHS) provides technical assistance for these surveys. The 2005-06 NFHS-3, the third in the series of these national surveys, was preceded by NFHS-1 in 1992-93 and NFHS-2 in 1998-99. Like NFHS-1 and NFHS-2, NFHS-3 was designed to provide estimates of important indicators on family welfare, maternal and child health, and nutrition.

NFHS-3 used three types of questionnaires: the Household Questionnaire, the Women's Questionnaire, and the Men's Questionnaire. The overall content and format of the questionnaires were determined through a series of workshops and meetings held in 2005-06.

The *Household Questionnaire* was used to list all usual residents in each sample household plus any visitors who stayed in the household the night before the interview. For each person listed, information was collected on age, sex, marital status, relationship to the head of the household, and education. For children age 0-4 years, information was collected on birth registration. Questions were asked about school/college attendance for children age 5-18 years, and questions were asked about the activities of children age 5-14 years. The Household Questionnaire also collected information on the main source of drinking water, type of toilet facility, source of lighting, type of cooking fuel, religion and caste/tribe of the household head, ownership of a house, ownership of agricultural land, ownership of livestock, ownership of other selected items, and whether the household had a BPL (Below Poverty Line) card. Information was also collected on health issues such as the prevalence of tuberculosis, use of private or public health facilities, and ownership of mosquito nets.

The *Women's Questionnaire* was employed to interview all women (ever-married and never-married) age 15-49 who were usual residents of the sample household or visitors who stayed in the sample household the night before the survey. The questionnaire covered topics like background characteristics, reproductive health and behaviour, marriage and cohabitation, knowledge and use of contraception, quality of care and contact with health personnel, antenatal and postnatal care etc. For the main purpose of this study, an additional section of NFHS-3 is being used. This section specifically asks questions about utilization of ICDS services. It is the first time that any specific set of questions on the utilization of this biggest running health and nutrition program were asked in a national level survey.

The *Men's Questionnaire* was employed to interview men age 15-54 who were usual residents of the sample household or visitors who stayed in the sample household the night before the survey. The Men's Questionnaire contains a subset of questions that are covered in the Women's Questionnaire, plus some additional questions only administered to men.

Sample Design: The urban and rural samples within each state were drawn separately and, to the extent possible, unless oversampling was required to permit separate estimates for urban slum and non-slum areas, the sample within each state was

allocated proportionally to the size of the state's urban and rural populations. A uniform sample design was adopted in all states. In each state, the rural sample was selected in two stages, with the selection of Primary Sampling Units (PSUs), which are villages, with probability proportional to population size (PPS) at the first stage, followed by the random selection of households within each PSU in the second stage. In urban areas, a three-stage procedure was followed. In the first stage, wards were selected with PPS sampling. In the next stage, one census enumeration block (CEB) was randomly selected from each sample ward. In the final stage, households were randomly selected within each selected CEB.

Sample Implementation and Data Collection: NFHS-3 fieldwork was carried out in two phases, in order to achieve better coordination and supervision in the implementation of the survey. Twelve states were canvassed in the first phase and the remaining 17 states were canvassed in the second phase. First-phase data collection was carried out from November 2005 to May 2006. Second-phase data collection was carried out from April to August 2006.

5.3.2 Data in Present Study

The data has been provided by DHS as distribution files where data from the three questionnaires has been brought down to a lower analytic level. For the purpose our analysis, we use the All Birth (BR) file.

Birth Recode files contain an entire woman's questionnaire for each child listed in her birth history. Sections that repeat in the questionnaire, such as the birth history, neonatal/antenatal care, and immunization/ nutrition, only occur once, for the child in question. If the child was not born in the past 5 years, these latter sections will be blank. A maximum of 20 children for each woman can be listed. The woman (mother) will have a de facto residency status, but her child may or may not be, and in fact may not even be a resident of the household. The child is the case, with the identifier being created from the household identification number and the child's birth order, the youngest child being listed first.

The study sample used in this analysis consists of children up to five years/ 59 months of age. Since almost 80 percent of ICDS projects are running in rural areas, we restrict our analysis to rural areas only. We have a total of around 26000 children, depending

on the dependent variable we use. We analyse 26065 children for weight-for-age, out of which there are 13491 boys and 12574 girls. For height-for-age dependent variable, there are 26038 children out of which 13482 are boys and 12556 are girls.

Dependent Variables

Anthropometric Measures: The health status (defined as H in equation 5) can be captured by different variables, which are typically either self-reported, subjective measures, or objective measures such as height, weight, or body mass index (Falkner and Tanner 1986). We focus here on two commonly used measures of long-term nutritional status, height for age as well as weight for age (Trapp and Menken 2005). The anthropometric index of height-for-age (HAZ) reflects pre and post-natal growth, and deficits in height-for-age show the ''long-term, cumulative effects of inadequacies of nutrition and/or health'' (Gillespie & Haddad 2001). A child is classified as "stunted" if her or his HAZ is below –2 standard deviations from the median of the National Centre for Health Statistics/World Health Organization international growth reference (WHO 1995). A child's weight-for-age Z-score (WAZ), the most commonly used measure of child nutritional status, is a summary measure of weight-for-age. A child is considered to be "underweight" if WAZ is less than –2 standard deviations from the international reference (ACC/SCN 2000).Wasting is another measure of nutritional status defined as low weight-for-height, and is often associated with starvation and/or severe disease (Gillespie & Haddad 2001).

The NFHS-3 collected anthropometric measures and health and vaccination histories for all children aged 0–60 months whose mothers were surveyed. In order to estimate the effects of maternal education on child nutritional status, our sample from the NFHS is limited to children between the ages of 0–60 months for whom these anthropometric data are available.

Explanatory Variables

The list of predictors of child nutritional status that we use includes a series of variables measured at the child, household and village level. We exclude variables such as housing characteristics, labor supply and asset ownership (the NFHS does not include information on expenditure or income).

Children's Characteristics (Variables at the Children Level)

One of the child characteristics included in the model is child age. This is expressed in one-year bracket, with the base group being children 0-12 months of age in order to accommodate well known age-specific patterns in nutritional status (Shrimpton et al. 2001). The variable measuring children's ages is formulated as a step dummy, with children 0–1 year old (0–11 months) the reference category and indicator variables for children aged 1–2 years (12–23 months) , 2-3 years (24–35 months), A 0-1 gender dummy variable included in the model where "male child" is taken as the reference category. Further, the number of times a child is breastfed also enters into the model as an important explanatory variable.

Household Characteristics (Variables at the Household Level)

Mother's Characteristics: Age and age of mother at the first birth are included in the regression analysis as important characteristics of the mother. Younger mothers typically have higher risks of poor pregnancy outcomes. While age may reflect biological factors, it also reflects socioeconomic considerations including standing in the household hierarchy. According to the bargaining literature on household decisions status could influence those resources that the mother may receive for herself as well as for her child, possibly leading to adverse nutrition consequences (Smith et al. 2003). Child health is also affected by the decision-making power of women within the family. Women generally are the primary care givers in their home, devoting more time to the protection and care of their children than men. To construct a variable to define the relative power of mother within the household, we use the information about the role of mother in decisions regarding her own health. The NFHS data asks the respondent about "final say in her own health care". When the respondent decides alone, or with husband or any other family member (in rural India, the role of mother-in-law is important in decisions regarding a woman's health), it is taken that she has a say in her own health. A dummy variable measures her status in the family.

Education of the mother is treated as step dummy variables, with "no education" being the reference category and indicator dummies for primary, secondary and higher education.

ICDS program also aims to create and generate awareness among women, particularly pregnant women and adolescent girls; therefore we include variables that reflect this. Knowledge and awareness can be viewed independent of education level of the mother. Measures of mothers' knowledge concerning disease treatment and prevention, and nutritional requirements, are limited in the NFHS and several indirect measures are used to create an index of general health knowledge. Moreover, it would not be wrong to conclude that a woman who uses television, newspapers and other means of communication frequently is more likely to gain knowledge about health programs and problems from these channels. Mother's capacity to gather awareness and knowledge is important in determining her ability to make use of resources available as well as make to use programs. We use the following questions to construct knowledge variables: "Ever participated in a literacy program, frequency of reading newspaper, frequency of listening to radio and frequency of watching television". We create dummy variables for each of these, where presence of the attribute is equal to 1.

Socio-Economic Variables: Among household characteristics, religion is an important factor in determining health of a child and use of program and medical services. The reference category is being a Hindu, while dummies indicate Muslim, Christian and Sikh. Another variable we include is that of caste of head of household (and therefore of household). The reference category is belonging to SC/ST. Sex of head of household is also included as an explanatory variable.

The household characteristics taken into account are household size, sex of the head of the of household, water and toilet use, location in an urban or rural area, economic status, and country of residence. Household size is the number of people who usually live and eat together.

Households' health environments are measured using indicators of type of water and latrine use. For water, the reference category is the use of surface water, and dummy variables for well and piped water indicate safer water use. For latrine use, the reference category is of no latrine, while dummy for community owned toilet. For higher degree of sanitation, the reference category is other latrine, while dummy variable is for flush latrine.

The NFHS-3 data does not provide any measure of income of households. However, household wealth is a good approximation of income. NFHS-3 data provides wealth

index which was constructed form the household level data using Principal Components Analysis (PCA). The input information for this came from household ownership of items ranging from furniture and vehicles; to dwelling characteristics and to whether household had any bank or post office account. Each asset was assigned a weight (factor score) generated through PCA, and the resulting asset scores were standardized. The sum of the scores of the assets possessed by each household resulted in that household's wealth index factor. The data is presented in quintiles.

Table 5.1 presents a list of variables used in this analysis.

Table 5.1: Variables of the Study

Variable	Type
Children's Nutritional Status	
Child's height-for-age Z-score	Continuous
Whether child is stunted (=1 if stunted; =0 otherwise)	Dichotomous
Child's weight-for-age Z-score	Continuous
Whether child is underweight (=1 if underweight; =0 otherwise)	Dichotomous
Proximal Determinants	
Woman's Nutritional Status	
Woman's body mass index	Continuous
Whether woman is underweight (=1 if underweight; =0 otherwise)	Dichotomous
Child Characteristics	
Child aged 0–11(=1 if age between 0-11 months; =0 otherwise)	Dichotomous
Child aged 12–35 (=1 if age between 12-35 months; =0 otherwise)	Dichotomous
Child aged 36–59(=1 if age between 36-59 months; =0 otherwise)	Dichotomous
Child's sex (=1 if male; =0 if female)	Dichotomous
Characteristics of Woman and Partner	
Woman's age (in years)	Continuous
Woman's age at first birth	Continuous
Woman's education (=1 if illiterate; =0 otherwise)	Dichotomous
Woman's education (=1 if primary; =0 otherwise)	Dichotomous
Woman's education (=1 if secondary; =0 otherwise)	Dichotomous
Woman's education (=1 if higher; =0 otherwise)	Dichotomous
Number of years of education	Continuous
Woman's Status	
Whether woman has any say in healthcare (=1 if yes; =0 if no)	Dichotomous
Knowledge of Woman	
Whether participated in literacy program (=1 if yes; =0 if no)	Dichotomous
Whether listens to radio (=1 if yes; =0 if no)	Dichotomous
Whether reads newspaper (=1 if yes; =0 if no)	Dichotomous
Whether watched television (=1 if yes; =0 if no)	Dichotomous
Household Characteristics	
Religion	
Religion is Hindu (=1 if Hindu; =0 otherwise)	Dichotomous
Religion is Muslim (=1 if Muslim; =0 otherwise)	Dichotomous
Religion is Christian (=1 if Christian; =0 otherwise)	Dichotomous
Religion is Sikh (=1 if Sikh; =0 otherwise)	Dichotomous
Caste	
Caste is SC/ST (=1 if SC/ST; =0 otherwise)	Dichotomous
Agricultural land (in acres)	Continuous
Household size	Continuous
Number of children under age 5	Continuous
Well water used (=1 if well water used; =0 if other sources)	Dichotomous
Piped water used (=1 if Piped water used; =0 if other sources)	Dichotomous
No latrine used (=1 if No latrine; =0 if yes)	Dichotomous
Pit latrine used (=1 if well water used; = if other sources)	Dichotomous
Flush latrine used (=1 if well water used; = if other sources)	Dichotomous
Wealth index	Continuous

Note: We test for all these variables, however only the ones that show significance and are in the best fit model are presented in estimation results.