OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 74, 5 (2012) 0305-9049
doi: 10.1111/j.1468-0084.2011.00670.x

Is there Complete, Partial, or No Recovery from Childhood Malnutrition? – Empirical Evidence from Indonesia*

Subha Mani

Department of Economics, Fordham University, 441 East Fordham Road, Bronx, New York 10458 (e-mail: smani@fordham.edu)

Abstract

This article uses a dynamic panel data model to identify the impact of early nutritional deficiencies on individuals' health status in later ages. We find that poor nutrition at young ages causes some, but not severe retardation in the growth of future height indicating partial recovery from chronic malnourishment. The results also indicate that – younger children, stunted children, and children who live in communities with six or more health posts exhibit larger recovery. The estimation strategy used here is especially attractive as it relies on weaker stochastic assumptions compared to earlier work in the literature.

I. Introduction

Research shows that early nutritional deficiencies as measured by anthropometric outcomes are associated with fewer grades of schooling, impaired cognitive development, and lower height attained during later ages. These factors further combined together affect lifetime earnings and productivity.[1] However, if children are able to recover from deficits in health status caused by early nutritional deficiencies, then some of the negative consequences associated with poor nutrition may be mitigated early on. The main objective of this article is to identify the extent to which malnourished children can correct for some of the poor nutritional outcomes from their past.

We estimate a dynamic conditional health demand function where health status from the current period is expressed as a function of lagged health status and current period prices, parental characteristics, and community resources (Hoddinott and Kinsey, 2001; Fedorov

*I thank seminar participants at Fordham University, Monash University, Northeast Universities Development Consortium (NEUDC) meetings, PAA meetings, Rutgers University, University of Melbourne, University of Southern California for helpful comments and suggestions. I am thankful to Utteeyo Dasgupta, Rahul Giri, John Ham, John Hoddinott, Jeffrey B. Nugent and Duncan Thomas for helpful comments and advice. I am grateful to the editor and the anonymous referee for providing useful suggestions. I am especially indebted to John Strauss for his guidance, continuous encouragement and support. Financial support was provided by the College of Letters, Arts, and Sciences, University of Southern California. All remaining errors are mine.
JEL Classification numbers: I10, O12, R20, D10.

[1]See Glewwe and Miguel (2008) and Strauss and Thomas (2008) for recent reviews.

and Sahn, 2005; Alderman, Hoddinott and Kinsey, 2006). The coefficient on the lagged health status captures the extent of recovery from childhood malnutrition, also known as 'catch-up'.

A coefficient of zero on the lagged health status indicates 'complete catch-up', that is, malnourished children who suffer from chronic nutritional deficiencies at a young age will not experience permanent growth retardation. A coefficient of one on the lagged health status indicates 'no catch-up', that is, malnourishment during childhood results in permanent growth faltering, affecting long-run physical well being. A coefficient between zero and one indicates 'partial catch-up', that is, children are able to recover from some, but not, all of the deficits in health status caused by early nutritional deficiencies (Hoddinott and Kinsey, 2001; Fedorov and Sahn, 2005; Alderman *et al*., 2006).

The major difficulty in estimating the dynamic relationship comes from the presence of unobservables such as child's innate ability to fight diseases, parental preferences and community connections that are likely to be correlated with the one-period lagged health status. The presence of random measurement error in anthropometric outcomes also makes it difficult to obtain an unbiased estimate of the catch-up coefficient.

Many of these econometric concerns do not arise with the use of experimental data, as collected by the INCAP (Institute of Nutrition of Central America and Panama) study (Habicht, Martorell and Rivera, 1995; Martorell, 1995; Martorell, 1999). However, it is both difficult and expensive to obtain such experimental data. In the absence of such experimental data, studies have used dynamic models to identify the extent to which childhood malnutrition affects subsequent health status (Johnston and Macvean, 1995; Adair, 1999; Hoddinott and Kinsey, 2001; Fedorov and Sahn, 2005; Alderman *et al.*, 2006). Among these, Adair (1999) and Johnston and Macvean (1995) fail to address econometric problems associated with sample attrition and endogeneity of the lagged dependent variable biasing their catch-up estimate.

Fedorov and Sahn (2005) use the Arellano and Bond (1991) and Arellano and Bover (1995) type estimation strategies that rely on the assumption of lack of serial correlation in the error terms; usually not satisfied in dynamic panel data models (Deaton, 1997; Blundell and Bond, 1998; Blundell, Bond and Windmeijer, 2000). The two-stage least squares (2SLS) method adopted in Hoddinott and Kinsey (2001) addresses random measurement error bias but may not address the omitted variables bias problem.[2] On the other hand, the maternal fixed-effects (MFE) estimation strategy adopted by Hoddinott and Kinsey (2001) addresses the omitted variable bias problem but does not address the measurement error bias.

Alderman *et al.* (2006) use the maternal fixed-effects instrumental variable (MFE-IV) estimation strategy, which addresses biases coming from measurement error in data and other household and community specific time-invariant unobservables. However, individual-specific time-invariant unobservables such as the child's innate ability to fight diseases are treated as random.[3]

This article addresses these methodological constraints and contributes to the existing literature in many ways. First, the article brings out the extent to which early nutritional deficiencies affect health status in later ages. Second, to our knowledge the article is the first

[2] See Mani (2008) and Strauss and Thomas (2008) for discussion.

[3] The child's genetic ability to fight diseases and absorb nutrients could potentially be correlated with the instruments used – no. of days the child was living prior to August 1980.

in this literature to examine if catch-up effects differ with age, duration, and community resources.[4] Third, the empirical strategy used in this article relies on weaker stochastic assumptions compared to earlier work in the literature. For instance, the preferred first-difference generalized method of moments (FD-GMM) estimation strategy used here addresses both the omitted variables bias and the measurement error bias using instruments that do not rely on assumptions such as – (i) lack of serial correlation in the error terms, (ii) the lack of correlation between the instruments and the time-invariant unobservables, and (iii) no measurement error in anthropometric data. Fourth, we bring out the pros and cons associated with using different IV estimation strategies in the context of a dynamic panel data model. Finally, the regression results are robust to econometric concerns such as weak instruments and sample attrition.

Indonesia provides an appropriate setting for this study. Despite achieving sustained economic growth in the early 1990s, more than 40% of children (below 5 years) suffered from chronic nutritional deficiencies that result in growth faltering (Strauss *et al*., 2004b; Mani, 2008). Stunting is a serious source of concern among policy makers in Indonesia. The extent to which these early nutritional deficiencies affect future well-being is an empirical question of interest to both policy makers and researchers. To address this question, a panel data set is constructed using observations on children between the age of 3 and 59 months in 1993, followed through the 1997 and 2000 waves of the Indonesian Family Life Survey (IFLS). Height in cm is a well-established indicator of chronic nutritional deficiency and long-term physical well-being, and therefore used as a suitable measure of health status in this article (Tanner, 1981; Martorell and Habicht, 1986; Strauss and Thomas, 1995; Martorell, 1999).

The FD-GMM estimator used here yields a catch-up coefficient of 0.23. A coefficient of 0.23 suggests partial catch-up effects; that is, malnutrition during childhood causes some but not significant growth retardation in future height. Using the FD-GMM estimation strategy, we also find that – (i) stunted children exhibit larger catch-up effects compared to children who do not suffer from growth faltering at an early age, (ii) younger children have larger catch-up potential than older children, (iii) children who live in communities with six or more health posts exhibit marginally larger catch-up effects, and (iv) catch-up potential also varies with the length of time between survey rounds (duration).

The article is organized as follows. Section II outlines the conceptual framework and section III describes the data used for estimation. Empirical findings are discussed in section IV and finally concluding remarks follow in section V.

II. Conceptual framework

A theoretical model of the determinants of health outcomes is outlined here as means for guiding the variables that appear as regressors in the empirical specification. Following Strauss and Thomas (2008), the household maximizes expected lifetime utility – U equation (1), subject to a lifetime budget constraint equation (2), and a period specific dynamic child health production function equation (3).[5]

[4]Fedorov and Sahn (2005) and Alderman *et al.* (2006) have previously examined age differential catch-up effects alone.

[5]See Strauss and Thomas (2008) for a similar, yet even more general framework of the model.

$$\text{Max}: U = E_t \sum_{t=0}^{T} \beta^t u_t(C_t, H_t, L_t; \theta_{pt}) \tag{1}$$

$$s.t: A_T = \prod_{t=0}^{T}(1+r_t)A_0 + \sum_{t=0}^{T}\prod_{\tau=t}^{T}(1+r_\tau)[w_t(T_t - L_t) + \pi_t - p_t^c C_t - p_t^m M_t] \tag{2}$$

$$H_t = f(H_{t-1}, M_t, I_t, D_t, \theta_{ct}, \theta_c, \mu_{ht}, \mu_h, G). \tag{3}$$

The theoretical framework assumes that: (i) household's lifetime utility function is additively separable over time (Deaton and Meullbauer, 1980; Fedorov and Sahn, 2005; Strauss and Thomas, 2008), (ii) the one-period lagged health status in equation (3) is assumed to be a sufficient statistic that captures the impact of all health inputs, environmental factors, and other time-varying characteristics starting from birth up until the last observed period in the sample (Strauss and Thomas, 1995, 2008; Hoddinott and Kinsey, 2001; Mani, 2008), (iii) the sub-utility functions are quasi-concave and twice differentiable, and (iv) the household can potentially borrow and or lend against its future in each period t.[6] The sub-utility function u_t depends upon food and non-food consumption goods, C_t, leisure, L_t, and health status of the child, H_t. The household's utility is also affected by unobserved preference shocks, θ_{pt}. The subjective discount factor is captured by β. The vector of prices of food and non-food consumption goods is captured by p_t^c. Similarly, p_t^m is a vector of price of health inputs, w_t is the wage rate (price of leisure), T_t is parents' total time endowment and A_0 stands for assets the households owns at the beginning of period 0. Profit income from farm and non-farm activities and all other sources of non-labour income is captured by π_t. The expectations operator, E_t is conditional on the information available at time t.

M_t captures health inputs used towards the maintenance and or improvement of child health, I_t characterizes the environment where the child lives capturing infrastructure availability and disease environment in the community, and D_t reflects all time-varying demographic characteristics such as the child's age. All time-varying health shocks like fever and diarrhoea are captured by θ_{ct}, while, θ_c summarizes information about all time-invariant characteristics such as the child's gender and time-invariant health endowments like the child's innate ability to absorb nutrients and fight diseases. Household specific time-varying and time-invariant demographics and background characteristics such as parents' rearing and caring practices are captured by μ_{ht} and μ_h respectively. Finally, G summarizes information about all genetic endowments capturing genotype and phenotype influences that affect child health.

The choice variable, M_t here will depend upon prices from many periods in an unrestricted way since the choice of health today affects health in all future periods. The optimal level of child health input M_t^*, determined by the household, can be written as follows:

$$M_t^* = f(H_{t-1}, I_t, P_t^c, P_t^m, w_t, \lambda, D_t, \pi_t, \theta_{ct}, \theta_c, \mu_{ht}, \mu_h, G, \theta_{pt}, E_t(Z_{t+j})) \quad \text{for } j = 1, \ldots, T-t \tag{4}$$

[6]We acknowledge that some of these assumptions such as additive separability are strong, but testing these assumptions is beyond the scope of this article.

and

$$Z = I_t, P_t^c, P_t^m, w_t, \lambda, D_t, \pi_t, \theta_{ct}, \theta_c, \mu_{ht}, \mu_h, G, \theta_{pt}.$$

Thus, M_t^* is a function of the one-period lagged health status, prices of consumption goods, prices of health inputs, wage rates, environmental factors, λ (marginal utility of wealth in period 0), a set of time-varying and time-invariant child level and household level characteristics, and household's expectations at date t about all future periods – prices, environmental characteristics, and household demographics as captured by the term Z. Empirically, we assume that the term Z enters the dynamic conditional health demand function only linearly.

The dynamic conditional health demand function eqaution (5) can be obtained by replacing M_t in equation (3) by M_t^*:

$$H_t^* = f(H_{t-1}, I_t, P_t^c, P_t^m, w_t, \lambda, D_t, \pi_t, \theta_{ct}, \theta_c, \mu_{ht}, \mu_h, G, \theta_{pt}, E_t(Z_{t+j})) \quad \text{for } j = 1, \ldots, T-t, \tag{5}$$

where

$$Z = I_t, P_t^c, P_t^m, w_t, \lambda, D_t, \pi_t, \theta_{ct}, \theta_c, \mu_{ht}, \mu_h, G, \theta_{pt}.$$

The empirical counterpart of the dynamic conditional health demand function can be written as follows:

$$H_{it} = \beta_0 + \beta_1 H_{it-1} + \sum_{j=2}^{R} \beta_j X_{jit} + \sum_{j=1}^{R} \delta_j Z_{ji} + \varepsilon_i + \varepsilon_h + \varepsilon_c + \varepsilon_{it}. \tag{6}$$

H_{it} and H_{it-1} are a child's height in cm measured at time t and $t-1$, respectively. All time-varying time-invariant regressors are captured by the X's and Z's respectively. A complete description of the variables included in the empirical specification is outlined in the data section.

There are four sources of unobservables in the dynamic specification – ε_i, ε_c, ε_h and ε_{it}. All time-invariant individual specific unobservables such as the child's inherent healthiness are captured by ε_i. Similarly, ε_h captures all time-invariant household specific unobservables like parental preferences and time discount rate, and ε_c captures time-invariant community specific unobservables like community endowments and political connections. Finally, ε_{it} includes child specific time-varying unobservables that are unknown to the econometrician at date t. The sequence of expected future household characteristics, prices, incomes and other factors affecting current health as captured by $E_t(Z_{t+j})$ are unknown to the econometrician at date t, and empirically enters the dynamic conditional health demand function through the time-varying error term (ε_{it}).

III. Data

The data used in this article comes from the 1993, 1997 and 2000 waves of the Indonesian Family Life Survey (IFLS). The IFLS is a large-scale longitudinal survey, the first wave of which was fielded during late 1993 and early 1994 (IFLS1). In IFLS1, 7224 households were interviewed. The first follow-up wave was surveyed during the second half of 1997 (IFLS2) just before the major economic and financial crisis in Indonesia. In IFLS2, 7629 households were interviewed of which 6,752 were original IFLS1 households and 877 were split-off households. The third wave (IFLS2+) was a special follow-up survey fielded during the late 1998. A 25% sub-sample of the original IFLS1 households was re-contacted in the late 1998 with the aim of analysing the immediate impact of the 1997–98 economic and financial crises. A total of 10,435 households were interviewed in 2000. Of these, 6,661 were original IFLS1 households and 3,774 households were split-off households. The sample surveyed in 1993–94 represented 83% of the Indonesian population living in 13 of Indonesia's 27 provinces at the time. The 13 provinces were selected to maximize representation of the population, capture the cultural socio-economic diversity of Indonesia, and yet be cost-effective given the size and the terrain of the country. A total of 321 enumeration areas (EAs)/communities were selected from these 13 provinces for final survey purposes.

The IFLS includes modules on measures of health, household composition, labour and non-labour income, schooling, fertility, consumption expenditure, immunization, prices, and community infrastructure allowing us to link household, individual and community level data both within a wave and over time (see Frankenberg and Karoly (1995), Frankenberg and Thomas (2000), Strauss *et al.* (2004a) for details on the IFLS).

The panel sample constructed here follows children between the age of 3 and 59 months in 1993 through the 1997 and 2000 waves of the survey. Our panel sample only follows children initially under 5 years since it is growth faltering during early life that largely determines the small stature exhibited by adults in developing countries (Martorell and Habicht, 1986; Satyanarayana *et al.*, 1989; Martorell, 1995). From IFLS1 complete information on age in months, gender, and height in cm is available for 2,203 children between the ages of 3 and 59 months. Of these 2,203 children in 1993, 1966 children could be followed in 1997, and 2,051 of the original IFLS1 sample could be re-contacted in 2000. A total of 1,819 children between the ages of 3 and 59 months in 1993 could be followed through the 1997 and 2000 waves of the IFLS – this sample excludes observations deleted due to measurement error in height attainments or age in months. Sample attrition between 1993 and 1997 is 10.76% and between 1997 and 2000 is 6.90%. Re-contact rates were much lower in 1997 compared to 2000.[7] A simple mean test on the difference in height attainments between all children in 1993 and children who were lost over time is -0.76 cm with standard error of 0.80. This difference is not statistically significant and indicates that attrition rates are not likely to be related to differences in initial period health status.[8]

The outcome variable of interest is height in cm. A nurse or a recently trained doctor visited each household to record various measures of physical health including height. The estimated specification includes individual level covariates such as lagged age in months (age in months at time $t-1$), male dummy (equals 1 if male, 0 if female) and interactions between the male dummy and lagged age in months to control for age and sex specific differences in growth in height. We also include for duration, where duration is the length of period measured in months between the two consecutive survey rounds and controls for the uneven gap between the three survey rounds. We control for mother's height in cm

[7]Thomas, Frankenberg and Smith (2001) find a similar pattern, where they observe higher household level attrition between 1993 and 1997 compared to 1997 and 1998. They attribute this decline in attrition rate to be associated with learning by doing in running a large-scale household level survey.

[8]Attrition related selection bias is further discussed in the robustness section of the article. Household level attrition is also fairly low in the IFLS, 1.4% per year between 1993 and 1997 and at about 1.3% per year between 1993 and 2000.

and father's height in cm to capture differences in genetic endowments. Mother's completed grades of schooling and father's completed grades of schooling are also included in the regressions to capture parental rearing and caring practices that affect the choice of health inputs. The individual level regressors can all be constructed from the household questionnaires.

Geographic information is available at four administrative levels in Indonesia (from smallest to the largest): community, kecamatan (subdistrict), kabupatan (district) and province. There are two problems associated with using the original community codes for constructing our community level covariates. First, community level data is only available for respondents residing in the 321 original IFLS communities. The IFLS does not provide detailed community level information for mover households except in 2000 (see Strauss *et al.* (2004a) for details). Therefore, we will not be able to control for community level time-varying observables for children who belong to the mover households in the years 1997 and in 2000. Second, we need to aggregate the data for estimating community fixed-effect models, because in 1993, the IFLS did not take anthropometric measurements for all children in the household. The 1993 survey only took anthropometric measurements of children who were selected for the detailed interviews and any other child (<6 years) who may have been present at the time of the interview. Consequently, there are communities in 1993, where, we have five or fewer children with complete anthropometric details.

In order to obtain the community level information, and estimate community fixed-effects models, the following decision rule is used to create the 'location' variable used here. The 'location' variable created here is assigned with the community code if there are five or more children residing in the same community. In cases where this criterion fails, the 'location' variable is assigned the code corresponding to the next level of aggregation, i.e. the kecamatan code following the same rule. Similarly the kabupatan and lastly (and occasionally) the province codes are assigned.[9] All community level characteristics reported for the years 1997 and 2000 vary at the location level.[10]

The aggregation method used to construct the community characteristics for the 1997 and 2000 waves could lead to major misrepresentation of the community resources.[11] To check for the extent to which this affects our catch-up estimates, we run a series of robustness checks, which are discussed later in the article.

Since we estimate a dynamic panel data model, the aggregation method used to construct the 'location' variable is only relevant for the community data used in 1997 and 2000. The community data from 1993 is used only to construct the instruments and we can use this data as available at the smallest administrative unit (desa/community) for the year 1993 and hence the data used to construct the instruments do not suffer from the aggregation problem that potentially exists for the 1997 and 2000 data.

As outlined in the theoretical framework, health status of the child also depends upon prices of – food and non-food consumption goods, leisure, and health inputs. In our regres-

[9]The kecamatan codes can be easily linked to other nationally representation data like the SUSENAS. The definition of a kecamatan and a kabupatan continues to change over time. In order to use systematic codes of the kecamatan and kabupatans over time, we use the 1999 BPS codes that define the kecamatan and kabuptan codes for all IFLS communities from all 3 years of the survey.

[10]The percentage of observations aggregated to a higher administrative unit is reported in Appendix, Table A1.

[11]We are grateful to the referee for raising this point and providing useful suggestions.

sions, we include for prices of three food consumption goods – rice, oil and condensed milk. The price data come from the community questionnaires and is calculated at the location level. Location level male and female wage rates are included in the regressions to account for the price of leisure. These are median wage rates calculated at the location level using wage data from the household questionnaire. To control for price of health inputs, we include distance to health centre (puskesmas) measured in km. Health centres are the most widely used public outpatient facility in Indonesia. The empirical specification also includes a rural dummy (equals 1 if rural, 0 if urban) to capture rural-urban differences in height attainments.

At the location level, we also control for three measures of community infrastructure. Our first measure, the percentage of households with electricity in the community captures availability and usage of electricity within a community. The second measure, a dummy variable that takes a value 1 if the community has paved road and 0 otherwise captures access to roads and markets in a community. Finally, the number of health posts in the community captures availability of health infrastructure in the community.

In the conceptual framework section outlined earlier, height today also depends upon the marginal utility of wealth (λ) at time zero that enters the dynamic conditional health demand function through the budget constraint. Since, λ is time-invariant, it will get first-differenced from our preferred specification. However, to treat λ as a constant we must rely on the assumption that households can freely borrow and lend in each period t. In reality most households in developing countries have access to imperfect credit markets constraining household resources. To ensure that the findings of this article are not sensitive to this assumption, we capture borrowing constraints using a lagged measure of log of households real per capita consumption expenditure (PCE) as an additional explanatory variable. We find that our results on the catch-up coefficient as reported in the results section are robust to treating λ as time-invariant.

Table 1 provides sample average and standard deviation of the dependent and independent variables used in the empirical specification.

IV. Results

Partial catch-up effects

The ordinary least square (OLS) estimate of lagged height is 0.52 (column 1, Table 2), indicating less than partial catch-up in height attainment. The condition of zero correlation between the error term and H_{it-1} may never be satisfied in dynamic models (Deaton, 1997; Blundell and Bond, 1998; Wooldridge, 2002). Hence with H_{it-1} endogenous, the standard OLS estimate of β_1 is likely to be biased and inconsistent.

The first IV strategy followed is the two-stage least-square (2SLS) estimation strategy with province fixed-effects, where the dynamic levels specification (equation (6)) is estimated using community level characteristics from 1993 as instruments for lagged height under the assumption that the community characteristics are exogenous.[12] Under this assumption the 2SLS estimation strategy addresses random measurement error bias

[12] We include for province fixed-effects and not location fixed-effects due to problems of multicollinearity that arise from little over-time variation in the location varying characteristics.

TABLE 1

Summary statistics

Variables	*Mean*	*Standard deviation*
Height (in cm)	105.86	19.42
Mother's height (in cm)	150.54	5.11
Father's height (in cm)	161.38	5.36
Mother's schooling	5.96	3.93
Father's schooling	6.90	4.33
PCE (log of real per capita household consumption expenditure)	9.87	0.76
Distance to health centre (in km)	5.07	4.58
Electricity	76.68	26.92
Male wage rate (in logs)	6.56	0.52
Female wage rate (in logs)	6.19	0.85
Price of rice (in logs)	0.86	0.20
Price of condensed milk (in logs)	5.17	1.52
Price of oil (in logs)	1.74	0.43
Dummy for paved road	0.74	0.44
Number of health posts	6.67	4.73
No. of observations	5,457	

resulting in a catch-up coefficient of 0.63 (column 2, Table 2). However, non-random placement of the community resources is likely to create unobserved correlation between the time-invariant unobservables and instruments, biasing β_1 (Rosenzweig and Wolpin, 1986).[13]

To remove the time-invariant unobservables, the dynamic model can be estimated in first-differences:

$$\Delta H_{it} = \beta_1 \Delta H_{it-1} + \sum_{j=2}^{R} \beta_j \Delta X_{jit} + \Delta\varepsilon_{it}. \tag{7}$$

OLS in first-differences removes the omitted variables bias and results in a catch-up coefficient of −0.18 (column 3, Table 2). However, it creates a larger downward bias compared to OLS in levels, magnifying the measurement error bias (Griliches and Hausman, 1986).

The second IV strategy followed here is an Arellano–Bond estimator, which uses community characteristics from 1993 and height in cm from 1993 as instruments for the FD one-period lagged height. The Arellano–Bond estimator yields consistent estimates under the assumption of lack of serial correlation in the error terms and exogeneity of the community characteristics. The Arellano–Bond strategy yields a catch-up coefficient of −0.09 (column 4, Table 2). It is shown later that the assumption of lack of serial correlation in

[13]Ghuman *et al.* (2005) show how the community level time-invariant unobservables could also be potentially correlated with other household specific observables creating an upward bias in the estimated coefficient of the household characteristics. The province fixed-effects used addresses some, but not all of the concerns associated with endogenous program placement.

TABLE 2

Dynamic child health demand function

RHS variables	(1) *OLS height*	(2) *Two-stage least squares height*	(3) *OLS in first-differences height*	(4) *Arellano–Bond height*	(5) *First-difference GMM height preferred estimates*	(6) *First-difference GMM height preferred estimates*
Lagged height/catch-up coefficient	0.52***	0.63***	−0.18***	−0.09**	0.23**	0.26**
	(0.02)	(0.18)	(0.03)	(0.04)	(0.09)	(0.12)
Male dummy	10.31***	10.17***				
	(3.43)	(3.33)				
Lagged age in months	0.46***	0.44***	0.41***	0.40***	0.43***	0.41***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)
Lagged age in months *male dummy	−0.16***	−0.15***	−0.18***	−0.16***	−0.17***	−0.13**
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)
Duration	0.79***	0.88***	0.17**	0.20**	0.48***	0.42***
	(0.06)	(0.14)	(0.08)	(0.08)	(0.11)	(0.14)
Duration * male dummy	−0.20***	−0.20***	−0.15	−0.14	−0.16	−0.09
	(0.07)	(0.07)	(0.10)	(0.10)	(0.12)	(0.13)
Duration * lagged age in months	−0.0075***	−0.008***	0.002***	0.001	−0.004**	−0.004**
	(0.00)	(0.002)	(0.0007)	(0.0008)	(0.001)	(0.001)
Duration * lagged age in months* male dummy	0.003**	0.002**	0.004***	0.003	0.003***	0.003***
	(0.001)	(0.001)	(0.0008)	(0.0008)	(0.0009)	(0.001)
Mother's height	0.19***	0.16***				
	(0.01)	(0.03)				
Father's height	0.13***	0.10***				
	(0.01)	(0.03)				
Mother's schooling	0.011	0.03				
	(0.02)	(0.03)				
Father's schooling	0.018	0.01				
	(0.02)	(0.02)				

TABLE 2

(*Continued*)

RHS variables	(1) *OLS height*	(2) *Two-stage least squares height*	(3) *OLS in first-differences height*	(4) *Arellano–Bond height*	(5) *First-difference GMM height preferred estimates*	(6) *First-difference GMM height preferred estimates*
Lagged PCE	0.53***	0.53***	−0.01	0.06	0.16	0.12
	(0.12)	(0.14)	(0.11)	(0.11)	(0.13)	(0.16)
Price of rice	0.52	1.09	−0.44	−0.53	0.16	
	(1.2)	(1.08)	(0.66)	(0.65)	(0.78)	
Price of oil	−0.11	−0.36	0.09	0.03	−0.07	
	(0.28)	(0.25)	(0.16)	(0.16)	(0.19)	
Price of condensed milk	−0.02	−0.05	−0.03	−0.0009	−0.014	
	(0.09)	(0.08)	(0.06)	(0.06)	(0.07)	
Rural dummy	−0.53	0.81	−0.84	−1.20	0.39	
	(1.23)	(0.93)	(0.99)	(1.00)	(1.27)	
Rural dummy * rice price	0.05	−1.47**	0.21	0.50	−0.55	
	(1.3)	(1.03)	(0.77)	(0.77)	(0.91)	
Health posts	−0.03	0.02	−0.01	−0.009	−0.005	
	(0.06)	(0.02)	(0.01)	(0.01)	(0.02)	
Male wage rate	0.30	0.49*	−0.13	−0.09	0.04	
	(0.29)	(0.25)	(0.16)	(0.16)	(0.21)	
Female wage rate	0.25	−0.05	0.04	0.13	0.11	
	(0.19)	(0.11)	(0.12)	(0.11)	(0.13)	
Distance to health centre	−0.02	−0.014	0.013	0.010	−0.016	
	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	
Dummy for paved road	0.0009	0.17	−0.03	−0.00008	−0.098	
	(0.33)	(0.22)	(0.25)	(0.25)	(0.27)	
Electricity	−0.001	−0.002	−0.007	−0.005	−0.002	
	(0.007)	(0.004)	(0.005)	(0.005)	(0.005)	
Location fixed-effects	Yes	No	No	No	No	No
Location interacted time dummies	No	No	No	No	No	Yes
Province fixed-effects	No	Yes	No	No	No	No

TABLE 2

(*Continued*)

RHS variables	(1) *OLS height*	(2) *Two-stage least squares height*	(3) *OLS in first-differences height*	(4) *Arellano–Bond height*	(5) *First-difference GMM height preferred estimates*	(6) *First-difference GMM height preferred estimates*
F-statistic on the excluded instruments		7.09		33.55	19.19	12.87
Hansen J statistic		0.49		7.44	2.89	0.66
		(0.48)		(0.05)	(0.40)	(0.71)
C statistic on PCE in 1993					1.28	0.47
					(0.25)	(0.49)

Notes: OLS, ordinary least square.
(i) ***significant at 1%, **significant at 5%, *significant at 10%.
(ii) Robust standard errors reported in column (1).
(iii) In column 2, robust standard errors clustered at the location level. Community characteristics from 1993 are used as IV's. Instruments included – electricity in 1993, dummy for paved road in 1993.
(iv) In column 3, OLS in first-differences with robust standard errors clustered at the location level.
(v) In column 4, robust standard errors clustered at the location level. Instruments used – no. of health posts in 1993, no. of health posts in 1993 interacted with age in 1993, age in 1993, height in 1993.
(vi) In column 5, robust standard errors clustered at the location level. Instruments used – no. of health posts in 1993, no. of health posts in 1993 interacted with age in 1993, age in 1993 and per capita consumption expenditure (PCE) in 1993.
(vii) In column 6, location interacted time dummies replace the location level time-varying observables. Instruments used are PCE in 1993, no. of health posts in 1993 interacted with age in 1993, and age in 1993.
(viii) P-values are reported for the C-statistic and the Hansen J-statistic.
(ix) Location is the level at which community resources are available for the years 1997 and 2000. While for 1993, the instrument variables are available at the smallest administrative unit.

the error terms is not satisfied here and hence an Arellano and Bond (1991) type estimator cannot be used to obtain an unbiased and consistent estimate on the catch-up term.[14]

The third IV estimation strategy followed is the FD-GMM estimation strategy which provides us with our preferred estimate on the catch-up term as it addresses both omitted variables bias (via first-differencing) and measurement error bias (via instrumental-variable techniques) in data. Two sets of preferred FD-GMM estimates are reported in columns 5 and 6 of Table 2. In column 5, we control for actual location level time-varying observables, whereas in column 6, we replace the location level time-varying observables with location interacted time dummies. In column 6, we make the catch-up estimate robust to the inclusion of all empirically observable and unobservable time-varying location characteristics. The FD-GMM estimates reported in columns 5 and 6 of Table 2 both use community characteristics and household characteristics from 1993 as instruments for the first-differenced lagged height.

The FD-GMM estimate reported in column 5, Table 2 use no. of health posts in 1993, PCE in 1993, age in 1993 and interactions between no. of health posts in 1993 with age in 1993 as excluded instruments for the first-differenced lagged height (see first-stage regression results in Appendix, Table A2). The FD lagged height results in a catch-up estimate of 0.23 (column 5, Table 2) indicating partial catch-up effects. In column 6, the first-differenced lagged height is identified using age in 1993, PCE in 1993, interaction between no. of health posts in 1993 with age in 1993 (see first-stage regression results in Appendix, Table A2). The specification in column 6, Table 2 results in a catch-up coefficient of 0.26, indicating partial catch-up.[15]

We argue that both the excluded instruments used in our preferred specifications reported in columns 5 and 6 of Table 2 are – (i) strongly correlated with the endogenous regressor, and (ii) uncorrelated with the error term in the second-stage regressions. First, the IVs used here are correlated with the endogenous regressor. Improvements in household PCE are also associated with improvements in child health and hence PCE in 1993 is likely to be strongly correlated with the first-differenced lagged height (Fedorov and Sahn, 2005; Ghuman *et al.*, 2005; Mani, 2008). Health posts in a community actively contribute towards meeting the health care needs of children and can be used to explain the subsequent changes in child health between 1993 and 1997. Health posts provide basic maternal and child health care to neighbourhood groups. They are primarily targeted towards meeting the health care needs of children between 0 and 5 years. Health posts provide immunization services, oral rehydration solution packets, and vitamin supplements on a monthly basis (Frankenberg, 2004). Health posts also provide food supplements to young children. In addition interaction terms between health posts in 1993 and child's age in 1993 capture age specific returns to availability of health post. While allowing for the interactive effects, we also control for age in 1993, which is exogenous.

Second, the IVs used here are likely to be uncorrelated with the error term in the second stage regressions. In a dynamic model, community characteristics from 1993 and PCE from 1993 have no direct impact on current health status except via their impact on the one-period lagged health status. Empirically, two-period lagged community characteristics

[14]The Arellano and Bover (1995) and the System GMM (Blundell and Bond, 1998) estimators also rely on the assumption of lack of serial correlation in the error terms (see Blundell *et al.*, 2000).

[15]The catch-up estimates reported in columns 5 and 6 are not statistically significantly different from each other.

could potentially be correlated with time-invariant unobservables due to endogenous program placement effects. At the same time, PCE from 1993 is also likely to be correlated with time-invariant unobservables such as parental preferences towards child health. However, conditional upon first-differencing and under the assumption that the random measurement error in community characteristics and PCE from 1993 is uncorrelated with the random measurement error in FD-lagged height; we can treat two-period lagged community characteristics and PCE as valid instruments for the first-differenced lagged height.[16]

Overall, both our preferred FD-GMM estimates reported in columns 5 and 6 of Table 2, indicate that childhood malnutrition causes some but not significant growth retardation in an individual's future physical well-being. The FD-GMM estimates reported in columns 5 and 6, address both the omitted variables bias and measurement error bias without relying on assumptions such as – (i) lack of serial correlation in the error terms, (ii) lack of correlation between the time-invariant unobservables and the instruments, and (iii) no measurement error in height attainment. Our coefficient estimate on the catch-up term relies on weaker stochastic assumptions compared to earlier work in the literature.

We find that a malnourished child in the absence of any catch-up would by adolescence, grow to be 4.15 cm shorter than a well-nourished child. However, in the presence of partial catch-up effects, i.e. a coefficient of 0.23 as estimated in this article indicates that a malnourished child, by adolescence will grow to be only 0.95 cm shorter than a well-nourished child.[17]

Our findings indicate that the catch-up coefficient varies significantly across the different estimation strategies followed and hence the validity of the underlying stochastic assumption becomes extremely important. Comparing our coefficient estimates with past research findings, we find that estimators such as the Arellano–Bond and Maternal fixed-effects estimators that address the omitted variables bias are likely to provide estimates that reflect larger catch-up potential than estimators like the 2SLS that addresses the measurement error bias problem alone. Hence, an estimator like the FD-GMM estimator used here that addresses both sources of biases relying on weaker stochastic assumption is likely to be preferred.

Limited role of child, household, and community characteristics

The coefficient estimates on individual and household level covariates are all largely consistent with that found in the literature (Hoddinott and Kinsey, 2001; Mani, 2008). First-differencing removes all time-invariant variation among the right hand side regressors and additional instrumenting of the first-difference specification results in a loss of variation over time. Both these factors explain for the little role played by community characteristics in determining current health status in the dynamic specification.

Catch-up effects differ with age, level of malnourishment, duration and community characteristics

It is usually hypothesized that younger children experience larger catch-up effects than older children (Martorell and Habicht, 1986; Habicht *et al*., 1995). For example, Habicht

[16]The statistical relevance of the instruments is discussed in detail in the robustness section.

[17]To obtain these numbers, we follow the methodology outlined in Alderman *et al.* (2006).

et al. (1995) show that the impact of the nutritional intervention program in rural Guatemala had the most significant impact on improving the stature of children less than 3 years of age. This article attempts to find similar support by adding an interaction term between lagged height and lagged age in months in the dynamic specification. A positive and significant coefficient estimate on the interaction term will indicate smaller catch-up among older children. The FD-GMM estimates reported in column 1, Table 3 yields a coefficient estimate of 0.0010 on the interaction term indicating age differential catch-up effects, i.e. older children experience smaller catch-up compared to younger children.

We also examine if catch-up effects are larger among stunted children. To address this question, we add an interaction term between lagged height and the stunted dummy where the stunted dummy takes a value 1 if the child's height for age z-score is less than -2 in 1993 and 0 otherwise. Our preferred FD-GMM estimator yields a coefficient estimate of -0.04 on the interaction term (see column 2, Table 3). A negative and significant coefficient estimate on this interaction term indicates larger catch-up among stunted children.

Catch-up potential is also likely to vary between survey rounds. To estimate differential catch-up effects for the periods 1993–1997 and 1997–2000, we estimate our preferred

TABLE 3

Catch-up effects differ by – age, nutritional status, and duration

RHS variables	*FD-GMM* (1) *Height*	*FD-GMM* (2) *Height*	*FD-GMM* (3) *Height*
Lagged height	0.24**	0.20**	0.70***
	(0.09)	(0.08)	(0.25)
Lagged height * lagged age in months	0.0010*		
	(0.0006)		
Lagged height * Stunted		–0.04***	
		(0.0009)	
Lagged height * duration			–0.014*
			(0.007)
Kleibergen–Paap rk Wald F-statistic	12.002	15.78	7.02
Hansen J-statistic	0.22	2.54	0.16
	(0.89)	(0.27)	(0.98)
C-statistic on per capita consumption expenditure (PCE) in 1993	0.07	0.64	0.12
	(0.78)	(0.42)	(0.72)

Notes: (i) ***significant at 1%, **significant at 5%, *significant at 10%.
(ii) In columns 1 and 2, robust standard errors clustered at the location level.
(iii) The variable stunted take a value 1 if the child has HAZ< -2 in 1993 and 0 otherwise.
(iv) In column 1, instruments used – PCE in 1993, no. of health posts in 1993, age in 1993, and age in 1993 interacted with no. of health posts in 1993.
(v) In column 2, instruments used – PCE in 1993, no. of health posts in 1993, no. of health posts in 1993 interacted with the variable stunted, PCE in 1993 interacted with the variable stunted.
(vi) In column 3, instruments used – PCE in 1993, no. of health posts in 1993, no. of health posts in 1993 interacted with age in 1993, age in 1993, PCE in 1993 interacted with duration.
(vii) All other covariates (RHS variables) are as included in column 5, table 2; these are suppressed and available from the author upon request.
(viii) P-values are reported for the C-statistic and the Hansen J-statistic.

TABLE 4

Catch-up effects vary with the no. of health posts in community

RHS variables	*FD-GMM* (1) *Height*	*FD-GMM* (2) *Height*
	Communities with less than 6 health posts	*Communities with more than or equal to 6 health posts*
Lagged height/catch-up coefficient	0.37*	0.23*
	(0.19)	(0.12)
Kleibergen–Paap rk Wald F-statistic on the excluded instruments	9.29	19.75
Hansen J-statistic	1.38	0.24
	(0.50)	(0.61)

Notes: (i) In columns 1 and 2 robust standard errors clustered at the location level.
(ii) ***significant at 1%, **significant at 5%, *significant at 10%.
(iii) In column 1, instruments used – no. of health posts in 1993, age in1993, and per capita consumption expenditure (PCE) in 1993.
(iv) In column 2, instruments used – PCE in 1993 and no. of health posts in 1993.
(v) All other covariates are as included in column 5 of Table 2; these are suppressed and available from the author upon request.
(vi) P-values are reported for the C-statistic and the Hansen J-statistic.

FD-GMM specification, now adding an interaction term between lagged height and duration (see column 3, Table 3). The negative sign on the interaction term indicates that the longer the duration the greater the catch-up. To compute the differential catch-up effects for the periods 1993–97 and 1997–2000, we evaluate the interaction term at 34 months (average duration during 1997–2000) and at 47 months (average duration during 1993–97). We obtain a catch-up coefficient of 0.12 for the longer duration (1993–97) and a catch-up coefficient of 0.25 for the shorter (1997–2000) duration, suggesting that the catch-up effect vary with duration.[18]

A policy question of interest that has still not been addressed in the literature is if certain specific policy intervention could enhance catch-up potential among malnourished children? In static models, improved access to health care at the community level has significant effects on improving children's long-run nutritional status (Frankenberg and Thomas, 2001). Can greater access to health care facility at the community level also enhance catch-up potential among children?

To examine this question we stratify our sample into two categories – children who have access to more than six health posts in the community and children who have access to less than six health posts in the community.[19] We use our preferred FD-GMM estimator to obtain catch-up effects for our stratified sample. The stratification of the sample is based on access to community characteristics in period t. The regression results reported in Table 4 indicate that children who have access to six or more health posts in the community exhibit larger catch-up effects compared to children who do not. These catch-up effects reported in

[18]We thank the referee for this suggestion.

[19]We also examined if catch-up effects differed by parental schooling and other community resources and do not find results to suggest differential catch-up effects among these factors. These results are available from the author upon request. In some cases, we also have severe weak instruments problem.

columns 1 and 2 of Table 4 are not statistically significantly different from each other. Yet, they provide some guidelines to policy makers on how improvements in access to health care facilities among children can result in larger catch-up among malnourished children and that investments in health care facilities have long-term policy implications.

Robustness

Instrument validity and sample attrition

The preferred estimates reported in this article are robust to important econometric concerns such as instrument validity and sample attrition. In the presence of weak correlation between the instruments and the endogenous regressors, the IV estimates reported here are likely to suffer from a larger bias and inconsistency compared to the bias obtained on the OLS parameter estimate (Murray, 2006). Recent work on weak IVs suggests the use of the Kleibergen–Paap Wald rk F-statistic is robust to the presence of heteroskedasticity, autocorrelation and clustering (Kleibergen and Paap, 2006). In the presence of a single endogenous regressor, the Kleibergen–Paap test statistic reduces to the First-stage F-statistic on the excluded instruments.

The first-stage F-statistic reported in columns 5 and 6 of Table 2 are 19.19 and 12.87. The F-statistic reported for our preferred estimates is greater than 10, satisfying the usual criteria for instrument validity, that is, our excluded IVs are strongly correlated with the endogenous regressor (Staiger and Stock, 1997).

It must also be the case that the instruments used are uncorrelated with the error term in the second stage regression. The Hansen J-statistic, a test of over-identifying restrictions is appended in Table 2. The Hansen J-statistic reported in Table 2 suggests that we cannot reject the null of over-identifying restrictions for specifications 5 and 6, but can still not conclude the validity of all the IVs. Murray (2006) points out that if all the IVs used in the over-identifying restrictions share a common economic rationale then even if one IV is invalid, it will cast doubt on all others. The over-identifying restrictions test is more meaningful when all IVs do not share the same economic rationale. For instance, in our case some of the IVs are lagged household characteristics while others are lagged community resources.

To test for the exogeneity of individual IVs, we use the C-statistic, where under the null, the suspect instrument is orthogonal to the error term. We test if PCE in 1993 can be treated as orthogonal to the error term in the second-stage regression for our preferred estimates reported in columns 5 and 6 of, Table 2. The C-statistic reported in Table 2 does not reject the null that the instrument is uncorrelated with the error term in the second-stage regression.[20]

We find similar estimates on our catch-up coefficient using alternative IVs which rely on similar stochastic assumptions enhancing the credibility of our IV estimates.[21]

A Hausman (1978) type test is also incorporated between the Arellano and Bond (1991) estimator and the first-difference GMM estimator. Under the null that there is no serial correlation in the error terms, the Arellano and Bond (1991) strategy must yield consistent

[20]We find similar evidence for PCE in 1993 for all other specifications reported in Tables 3 and 4.
[21]These are available from the author upon request.

and efficient parameter estimates on the first-differenced lagged height. However, if this assumption fails, the alternative first-difference GMM estimator must be chosen which is consistent and efficient under the alternative but not under the null. The first-difference GMM (in column 5, Table 2) estimator is tested against the Arellano and Bond (1991) (in column 4, Table 2) estimator. The estimated difference on the catch-up coefficients is 0.31 (standard error 0.07), rejecting the null of zero first-order and second-order serial correlation in the error terms.[22]

We also estimate our preferred FD-GMM specification without controlling for the first-differenced lagged PCE to test if the results on the catch-up coefficient are robust to the assumption that λ is treated as a constant. We find that the FD-GMM estimate results in a catch-up coefficient of 0.21, which is not statistically significantly different from our preferred estimate of 0.23.

The coefficient estimates on the catch-up term is also robust to sample attrition. Attrition can be a problem only if; either the observable or unobservables that result in attrition are correlated with the error term in the specification of interest (Fitzgerald, Gottschalk and Moffitt, 1998). Individual level attrition is also not a real concern in this paper, given the estimation strategy adopted here. First-differencing removes all potential sources of unobservables like the child's genetic endowments which are likely to be correlated with potential observables or unobservables that result in attrition. In the presence of first-differencing, the only possible remaining source of attrition is that arising from the presence of random health shocks, such as infectious diseases. These health shocks are likely to be uncorrelated with the health shocks in subsequent periods. Hence, attrition arising from the presence of random, time-varying health shocks is not likely to contaminate the parameter estimate on the lagged dependent variable.[23]

Construction of the community characteristics

The flexible definition used to construct the location variable (the geographic unit at which all community resources for 1997 and 2000 are constructed) could potentially misrepresent the true community resources available to the child and as a result affect our catch-up coefficient. To address this concern, we estimate two alternate specifications – (i) the first-difference GMM specification replacing the location interacted time dummies in column 6 of Table 2 with kecamatan (sub-district) interacted time dummies, and (ii) the first-difference GMM specification replacing the location interacted time dummies with kabu-

[22]A similar result emerges when we compare column 4 estimates with column 5, except that we have now dropped PCE in 1993 from specification 5.

[23]An OLS model on attrition is also estimated where the dependent variable, attrition is defined equal to 1 if the individual can be followed through the 1993, 1997 and 2000 waves of the IFLS, and zero otherwise. The right hand side regressors include height in cm, mother's schooling, father's schooling, mother's height, father's height, male dummy, age in months, log of real PCE, mother's age, father's age, rural dummy, and location indicators. All the right hand side regressors belong to the baseline survey year, 1993. The coefficient on height in cm from 1993 is 0.002 with a standard error of 0.004, indicating an insignificant impact on attrition. Among the other regressors mentioned above, it is only the rural dummy, which has a significant impact on attrition. Children residing in rural areas are more likely to be followed as compared to children residing in urban areas in the baseline year. This is similar to the findings reported in Thomas *et al.* (2001), where they find that household level attrition rates are higher in urban areas compared to rural areas. In summary, the OLS estimates verify that attrition is unrelated to endogenous observables like the child's health status from 1993 and measure of household income. Hence the parameter estimates reported in this paper are not likely to be confounded by selection issues. The complete regression results on attrition are available from the author upon request.

patan (district) interacted time dummies. The catch-up coefficients obtained in these two alternate specifications are 0.17 (std. error = 0.09) and 0.22 (std. error = 0.13), respectively. The community characteristics in 1993 are used as IVs in these alternate specifications and reflect the true community level resources available to the child. Hence we conclude that the catch-up coefficient is not sensitive to the geographic aggregation method used here.

The aggregation method used to construct the 'location' variable/community resources for the years 1997 and 2000 would especially be a concern if there is a lot of variation in the kind of resources available across communities (desa/village), the smallest geographic unit. To check this we run some analysis of variance association (ANOVA) to determine the extent to which the variation at the community level is arising from the variation across kecamatans (sub-district) and kabupatans (district). We run an ANOVA, where the unit of observation is at the 'village/desa' level and we regress health posts in 1997 on a set of kecamatan (sub-district) dummies and obtain an associated R-square of 0.96. This R-square tells us that 96% of the variation in health posts is across sub-districts and not within communities of a sub-district. Again for the same year, we similarly regress health posts in 1997 on kabupatan (district) dummies and obtain an R-square of 0.74, which tells us that 74% of the variation in health posts is across districts and not within communities of a district. We carry out a similar ANOVA exercise for 2000 and find that 92% of the variation in health posts is across sub-districts and about 76% of the variation is across districts. A similar ANOVA exercise is also performed for the other community variables such as – distance to health centre in km, percentage of households with electricity (electricity), dummy for the presence of paved road and prices of consumption goods. We find that in almost all cases over 90% of the variation in the community resources is across sub-districts and there is very little variation among communities within a sub-district. We find that around 70%–75% of the variation in the community resources is across districts and there is again smaller variation among communities within a district. Less variation among community resources within sub-districts and districts indicate lower possibilities of misrepresented resources through our aggregation method. Consequently, the geographic aggregations used here are not likely to affect the catch-up estimates reported in the paper.

To be certain that the community resources created using the 'aggregation' method do not confound our catch up estimate, we re-estimate our preferred specifications separately for the 'aggregated' sample and 'non-aggregated' sample. The aggregated sample includes observations on children for whom the community characteristics are aggregated to a higher level of geographic unit (sub-district, district or province) at least once between 1997 and 2000 and for the non-aggregated sample the community characteristics are available at the smallest geographic unit. We re-estimate our preferred first-difference GMM specifications for the aggregated and non-aggregated samples separately (see Appendix, Table A3). We find that the coefficient estimate on the catch-up term is not statistically significantly different between the aggregated and non-aggregated data. We can therefore conclude that our estimates on the catch-up coefficient are fairly robust to the aggregation method used in this paper.

Movers and non-movers

Selective migration can bias the estimated coefficient on the catch-up term. We find that about 21.22% of the children have moved at least once by the year 2000, of which only

11.76% moved to another community. For our purpose, a mover is one who has left his/her original community at least once between 1993 and 2000. Taking the average height over 3 years (1993, 1997 and 2000) we find that mover children are significantly taller than non-mover children raising concerns about sample selection (Appendix, Table A4). We check if our mover and non-mover children experience differential catch-up effects. Splitting the sample would normally create selection bias due to the presence of unobserved correlation between the time-invariant unobservables and migration. First-differencing though, allows us to address selection arising from the presence of time-invariant unobservables and allows us to estimate catch-up effects separately for both mover and non-mover children. The catch-up estimates for mover and non-mover children are reported in Appendix, Table A5. We find that mover children exhibit marginally larger (though not statistically significantly) catch-up than non-mover children. The catch-up estimates for the mover and non-mover sample are quite close to the estimates reported for the full sample in columns 5 and 6 of Table 2.

V. Conclusion

This article captures the extent to which childhood malnutrition affects subsequent health status. A dynamic conditional health demand function is estimated where the coefficient on the lagged dependent variable captures the extent of recovery, if any, from childhood malnutrition. A coefficient of 0.23 on the one-period lagged health status indicates partial catch-up in height attainments. In the presence of partial catch-up, by adolescence, a malnourished child will grow to be 0.95 cm shorter than a well-nourished child. In the absence of any catch-up, by adolescence, a malnourished child will grow to be 4.15 cm shorter than a well-nourished child.

There is only some evidence showing that catch-up effects are marginally larger among younger children than older cohorts. From a practical standpoint, the presence of partial catch-up effects and age-differential catch-up effects suggests that continued efforts must be made on the part of households and policy makers towards improving children's nutritional status at all ages. However, special emphasis must be on younger age groups, as their catch-up potential is still the largest.

We find that children who live in communities with six or more health posts exhibit larger catch-up effects. This result sheds light on some channels through which catch-up is possible.

The empirical techniques used here rely on weaker stochastic assumptions compared to earlier work, addressing both omitted variables bias and measurement error bias in data. The findings are also robust to common econometric concerns such as sample attrition and weak instrument.

Final Manuscript Received: July 2011

References

Adair, L. (1999). 'Filipino children exhibit catch-up growth from age 2 to 12 years', *Journal of Nutrition*, Vol. 129, pp. 1140–1148.

Alderman, H., Hoddinott, J. and Kinsey, B. (2006). 'Long-term consequences of early childhood malnutrition', *Oxford Economic Papers*, Vol. 58, pp. 450–474.

Arellano, B. and Bond, S. (1991). 'Some tests of specification for panel data: Monte Carlo evidence and an application of employment equations', *Review of Economic Studies*, Vol. 58, pp. 277–297.

Arellano, B. and Bover, O. (1995). 'Another look at the instrumental variable estimation of error-components models', *Journal of Econometrics*, Vol. 68, pp. 29–51.

Blundell, R. and Bond, S. (1998). 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of Econometrics*, Vol. 87, pp. 115–143.

Blundell, R., Bond, S. and Windmeijer, F. (2000). Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimators, Working Paper 00/12, The Institute for Fiscal Studies.

Deaton, A. (1997). *The Analysis of Household Surveys: A Micro Econometric Approach in Development Policy*, Johns Hopkins University Press, Baltimore, MD.

Deaton, A. and Muellbauer, J. (1980). *Economics and Consumer Behavior*, Cambridge University Press, New York.

Fedorov, L. and Sahn, D. (2005). 'Socioeconomic determinants of children's health in Russia: a longitudinal survey', *Economic Development and Cultural Change*, Vol. 53, pp. 479–500.

Fitzgerald, J., Gottschalk, P. and Moffitt, R. (1998). 'An analysis of sample attrition in panel data: the Michigan panel study of income dynamics', *Journal of Human Resources*, Vol. 33, pp. 251–299.

Frankenberg, E. (2004). Sometimes It Takes a Village: Collective Efficacy and Children's use of Preventive Health Care, Working Paper Series CCPR-028-04, California Center for Population Research.

Frankenberg, E. and Karoly, L. (1995). The 1993 Indonesian Family Life Survey: Overview and Field Report, DRU-1195/1-NICHD/AID, the RAND Corporation.

Frankenberg, E. and Thomas, D. (2000). The Indonesia Family Life Survey (IFLS): Study Design and Results from Waves 1 and 2, DRU-2238/1-NIA/NICHD, the RAND Corporation.

Frankenberg, E. and Thomas, D. (2001). 'Women's health and pregnancy outcomes: do services make a difference?' *Demography*, Vol. 38, pp. 253–265

Ghuman, S., Behrman, J., Borja, J., Gultiano, S. and King, E. (2005). 'Family background, service providers, and early childhood development in Philippines: proxies and interactions', *Economic Development and Cultural Change*, Vol. 54, pp. 129–164.

Glewwe, P. and Miguel, E. (2008). 'The impact of child health and nutrition on education in less developed countries', in Schultz T. and Strauss J. (eds), *Handbook of Development Economics*, Vol. 4, North Holland Press, Amsterdam, pp. 3562–3606.

Griliches, Z. and Hausman, J. (1986). 'Errors-in-variables in panel data', *Journal of Econometrics*, Vol. 31, pp. 93–118.

Habicht, J.-P., Martorell, R. and Rivera, J. (1995). 'Nutritional impact of supplementation in the INCAP longitudinal study: analytic strategies and inferences', *Journal of Nutrition*, Vol. 125, pp. 1042S–1050S.

Hausman, J. (1978). 'Specification tests in econometrics', *Econometrica*, Vol. 46, pp. 1251–1271.

Hoddinott, J. and Kinsey, B. (2001). 'Child growth in the time of drought', *Oxford Bulletin of Economics and Statistics*, Vol. 63, pp. 409–436.

Johnston, F. and Macvean, R. (1995). 'Growth-faltering and catch-up growth in relation to environmental change in children of disadvantaged community from Guatemala City', *American Journal of Human Biology*, Vol. 7, pp. 731–740.

Kleibergen, F. and Paap, R. (2006). 'Generalized reduced rank tests using the singular value decomposition', *Journal of Econometrics*, Vol. 133, pp. 97–126.

Mani, S. (2008). Is There Complete, Partial, or No Recovery from Childhood Malnutrition – Empirical Evidence from Indonesia, Fordham Discussion Series, Fordham University.

Martorell, R. (1995). 'Promoting healthy growth: rationale and benefits', in Pinstrup-Andersen P., Pelletier D. and Alderman H. (eds), *Child Growth and Nutrition in Developing Countries: Priorities for Action*, Cornell University Press, Ithaca, NY, pp. 15–31.

Martorell, R. (1999). 'The nature of child malnutrition and its long-term implications', *Food and Nutrition Bulletin*, Vol. 20, pp. 288–292.

Martorell, R. and Habicht, J.-P. (1986). 'Growth in early childhood in developing countries', in Falkner F. and Tanner J. (eds), *Human Growth, Methodology, Ecological, Genetic, and Nutritional Effects on Growth*, 2nd edn, Plenum, New York, pp. 241–262.

Murray, M. (2006). 'Avoiding invalid instruments and coping with weak instruments', *Journal of Economic Perspectives*, Vol. 20, pp. 111–132.

Rosenzweig, M. and Wolpin, K. (1986). *'Evaluating the effects of optimally distributed public programs'*, *American Economic Review*, Vol. 76, pp. 470–482.

Satyanarayana, K., Radhaiah, G., Mohan, K., Thimmayama, B., Rao, N., Rao, B. and Akella, S. (1989). 'The adolescent growth spurt of height among rural Indian boys in relation to childhood nutritional background: an 18 year longitudinal study', *Annals Human Biology*, Vol. 16, pp. 289–300.

Staiger, D. and Stock, J. (1997). 'Instrumental variables regression with weak instruments', *Econometrica*, Vol. 65, pp. 557–586.

Strauss, J. and Thomas, D. (1995). 'Human resources: empirical modeling of household and family decisions', in Behrman J. and Srinivasan T. (eds), *Handbook of Development Economics*, Vol. 3, Amsterdam: North Holland Press.

Strauss, J. and Thomas, D. (2008). 'Health over the life course, forthcoming' in Schultz T. and Strauss J. (eds), *Handbook of Development Economics*, Vol. 4, Amsterdam: North Holland Press.

Strauss, J., Beegle, K., Sikoki, B., Dwiyanto, A., Herawati, Y. and Witoelar, F. (2004a). The Third Wave of the Indonesian Family Life Survey (IFLS3): Overview and Field Report, WR-144/1-NIA/NICHD, the RAND Corporation.

Strauss, J., Beegle, K., Dwiyanto, A., Herawati, Y., Pattinasarany, D., Satriawan, E., Sikoki, B., Sukamdi and Witoelar, F. (2004b). *Indonesian Living Standards: Before and After the Financial Crisis*, the RAND Corporation, Santa Monica, CA.

Tanner, J. (1981). *A History of the Study of Human Growth*, Cambridge University Press, New York.

Thomas D., Frankenberg, E. and Smith, J. (2001). 'Lost but not forgotten: attrition and follow up in the Indonesian family life survey', *Journal of Human Resources*, Vol. 36, pp. 556–592.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge.

Appendix A

TABLE A1

Imputation of the location variable for the years 1997 *and* 2000

	1997		2000	
Distribution of the location variable	*Movers, as a % of the sample*	*Non-movers, as a % of the sample*	*Movers, as a % of the sample*	*Non-movers, as a % of the sample*
Community level aggregation	4.34	72.3	0.82	67.8
Kecamatan level aggregation	0.6	2.9	1.5	5.27
Kabupatan level aggregation	5.11	9.9	6.8	12
Province level aggregations	1.7	3	2.4	2.9

Notes: 1,819 observations in each year.

TABLE A2

First-stage regression results for the preferred estimates reported in columns 5 and 6 of Table 2

Excluded and included instruments from the first-stage regression	*First-stage regressions estimates for column 5, Table 2*	*First-stage regressions estimates for column 6, Table 2*
Included instruments		
First-differenced lagged per capita consumption expenditure (PCE)	−0.045 (0.19)	−0.05 (0.20)
First-differenced lagged age in months	0.49 ** (0.23)	0.45* (0.25)
First-differenced lagged age in months * male dummy	−0.03 (0.05)	−0.05 (0.06)
First-differenced duration	0.14 (0.29)	0.16 (0.33)
First-differenced duration * male dummy	−0.015 (0.14)	−0.05 (0.17)
First-differenced duration * lagged age in months	0.001 (0.003)	−0.0004 (0.003)
First-differenced duration * lagged age in months * male dummy	0.0009 (0.001)	0.0009 (0.001)
First-differences in rice price	−0.37 (0.88)	
First-differences in oil price	−0.08 (0.30)	
First-differences in condensed milk price	0.01 (0.08)	
First-differences in rural dummy	−1.00 (1.19)	
First-differences in rural dummy * price of rice	0.08 (0.84)	
First-differences in health posts	−0.014 (0.04)	
First-differences in male wage rate	−0.15 (0.30)	
First-differences in female wage rate	−0.03 (0.21)	
First-differences in distance to health centre	0.05** (0.02)	
First-differences in electricity	−0.005 (0.007)	
First-differences in dummy for paved road	0.10 (0.36)	

TABLE A2

(*Continued*)

Excluded and included instruments from the first-stage regression	*First-stage regressions estimates for column* 5, *Table* 2	*First-stage regressions estimates for column* 6, *Table* 2
Excluded Instruments		
No. of health posts in 1993	0.13*	
	(0.08)	
No. of health posts in 1993 * age in 1993	−0.002	0.001
	(0.002)	(0.0014)
PCE in 1993	1.17***	1.119***
	(0.26)	(0.31)
Age in 1993	−0.20***	−0.22***
	(0.04)	(0.05)

Notes: ***significant at 1%, **significant at 5%, *significant at 10%

TABLE A3

Catch-up estimates for aggregated vs. non-aggregated sample

Covariates	*FD-GMM aggregated height* (1)	*FD-GMM non-aggregated height* (2)
Lagged height	0.22*	0.23*
	(0.12)	(0.13)
F-statistic on the excluded instruments	7.82	10.98
Hansen J-statistic	1.00	2.37
	(0.80)	(0.49)
C-statistic on per capita consumption expenditure (PCE) in 1993	0.01	1.68
	(0.91)	(0.19)

Notes: (i) ***significant at 1%, **significant at 5%, *significant at 10%
(ii) Instruments used in columns 1 and 2 – PCE in 1993, no. of health posts in 1993, no. of health posts in 1993 interacted with age in 1993, age in 1993.
(iii) All other covariates (RHS variables) are as included in column 5, Table 2, but not reported.
(iv) P-values are reported for the C-statistic and the Hansen J-statistic.

TABLE A4

Height of mover and non-mover children over 1993–2000

	Mover	*Non-mover*	*Difference*
Average height	117.79	104.84	−12.94**
			(0.96)
Sample	428	5,029	

TABLE A5

Catch-up effects for mover and non-mover children

Covariates	*FD-GMM* (1) *height movers*	*FD-GMM* (2) *height non-movers*
Lagged height	0.29**	0.34**
	(0.13)	(0.15)
Kleibergen–Paap rk Wald F-statistic on the excluded instruments	9.08	7.15
Hansen J-statistic	6.5	0.18
	(0.08)	(0.91)
C-statistic on PCE in 1993	0.24	0.08
	(0.62)	(0.76)

Notes: (i) ***significant at 1%, **significant at 5%, *significant at 10%
(ii) In columns 1 and 2, robust standard errors clustered at the location level.
(iii) Instruments used in columns 1 – per capita consumption expenditure (PCE) in 1993, no. of health posts in 1993, no. of health posts in 1993 interacted with mother's schooling and no. of health posts in 1993 interacted with age in 1993.
(iv) Instruments used in column 2 – PCE in 1993, no. of health posts in 1993, no. of health posts in 1993 interacted with mother's schooling.
(v) P-values are reported for the C-statistic and the Hansen J-statistic.
(vi) All other covariates (RHS variables) are as included in column 5, table 2, but not reported.