The Economic Journal, **119** (*April*), 734–763. © The Author(s). Journal compilation © Royal Economic Society 2009. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

THE IMPACT OF IMPROVING NUTRITION DURING EARLY CHILDHOOD ON EDUCATION AMONG GUATEMALAN ADULTS*

John A. Maluccio, John Hoddinott, Jere R. Behrman, Reynaldo Martorell, Agnes R. Quisumbing and Aryeh D. Stein

Using a longitudinal survey from rural Guatemala, we examine the effect of an early childhood nutritional intervention on adult educational outcomes. An intent-to-treat model yields substantial effects of an experimental intervention that provided highly nutritious food supplements to children, a quarter century after it ended: increases of 1.2 grades completed for women and one quarter SD on standardised reading comprehension and non-verbal cognitive ability tests for both women and men. Two-stage least squares results that endogenise the actual supplement intakes corroborate these magnitudes. Improving the nutrient intakes of very young children can have substantial, long-term, educational consequences.

Throughout the world, there are hundreds of interventions that seek to improve the welfare of pre-school children through better nutrition, health, and psycho-social stimulation. In addition to their immediate effects, including improved survival and better child growth and development, investments in such interventions are often justified on the grounds that they provide longer-term benefits such as improved school readiness and educational attainment, as well as improved employment and health outcomes in adulthood. Yet only limited evidence exists to support claims regarding long-term impacts and many studies on this topic suffer from one or more

* This research was supported by National Institutes of Health grants TW-05598 on 'Early Nutrition, Human Capital and Economic Productivity' and HD-046125 on 'Education and Health Across the Life Course in Guatemala', National Institute of Child Health and Human Development grant 5R01HD045627-05 on 'Intergenerational Family Resource Allocation in Guatemala', and NSF/Economics grants SES 0136616 and SES 0211404 on 'Collaborative Research: Nutritional Investments in Children, Adult Human Capital and Adult Productivities'. We thank Alexis Murphy and Meng Wang for excellent research assistance in the preparation of the data for this article. We also thank participants for their comments during presentations at the 2003 Northeast Universities Development Consortium Annual Meetings, the 2003 Latin American and Caribbean Economic Association Annual Meetings, the 2003 Inter-American Development Bank Social Policy Network Conference, the 2004 Population Association of America Annual Meetings, the 2005 Minnesota International Economic Development Conference, the 2005 Econometric Society World Congress, the 2006 Mini-Conference on Development Economics in Quebec City, the World Bank, the Chinese University of Hong Kong, the Center for Global Development, Dartmouth College, George Washington University, University of California-Riverside, University College of London, University of Southern California, the Friedman School of Nutrition at Tufts, Wilfrid Laurier University and Williams College. Finally, we thank Jörn-Steffen Pischke and two anonymous referees for very useful comments on earlier versions of this article.

deficiencies in design, analyses, or interpretation.[1] First among these is the standard set of concerns applicable to all programme evaluations, including the comparability of treatment and control (or comparison) groups, the need for adequate sample sizes and the importance of accounting for selective attrition. Second is the need to use a sufficiently long time horizon to determine whether benefits fade out, persist, or even increase, over time – which may be especially important if some important benefits occur decades later. Third, typically only limited information is available on educational outcomes. For example, current school enrolment or completed grades are often used as the only indicator of educational attainment but these are only crude measures of what individuals know when they become adults, in part because both school quality and learning outside of school are likely to vary.

In this article, we assess the impact of one such intervention using methods and data that allow us to address these concerns. Nearly four decades ago, between 1969 and 1977, two nutritional supplements (a high protein-energy drink and a low-energy drink devoid of protein, randomly assigned at the village level) were provided to pre-school children in four villages in eastern Guatemala. Between 2002 and 2004, we traced and interviewed individuals who had been exposed to this intervention. Using these data, we estimate double-difference intent-to-treat models of exposure to the high-energy, high-protein intervention during the period from birth to 36 months of age on schooling attainment, reading comprehension, and non-verbal cognitive ability, all measured in adulthood. Our estimates control for family background characteristics – parental schooling, parental age, and a wealth index measured at the time of the intervention. Exploiting detailed historical studies undertaken in the survey areas over the past three decades, the estimates also control for relevant observed time-varying community-level factors that might otherwise adversely affect their accuracy and precision. We find significantly positive, and fairly substantial, effects of the nutritional intervention a quarter century later

[1] In the US, a small number of experimental evaluations of interventions focusing on pre-school children (and based on relatively small samples) – such as the Perry Pre-school Experiment, the Home Intervention Program for Pre-school Youngsters, and the Milwaukee and Abecedarian projects – find that such programmes generate higher grades of schooling attained, test scores, and incomes, as well as lower welfare participation rates and crime rates and out-of-wedlock births (Baker *et al.*, 1998; Ramey *et al.*, 1998; Schweinhart and Weikart, 1998). In addition, an important non-experimental evaluation of an intervention focusing on pre-school children (and based on a much larger sample), examined the Head Start programme. Estimates that control for mother and child-specific unobservable characteristics indicate that Head Start had positive effects on immunisation rates and test scores, and that it lowered grade repetition, primarily among whites and Hispanics (Currie and Thomas, 1995; Garces *et al.*, 2002). In developing country contexts, a body of literature, much of it outside economics, has explored the relationship between pre-school nutritional status and the education of school-age children and adolescents. There is evidence that undernourished children: score lower on tests of cognitive functioning; have poorer psychomotor development and fine motor skills; have lower activity levels; interact with others less frequently; delay initial school enrolment until older; have lower enrolment rates; and complete fewer grades of schooling (Johnston *et al.*, 1987; Behrman, 1996; Grantham-McGregor *et al.*, 1997; Grantham-McGregor *et al.*, 1999*a,b*; Alderman *et al.*, 2001*b*; Glewwe *et al.*, 2000; Glewwe and King, 2001; Behrman *et al.*, 2004; Alderman *et al.*, 2006; and Victora *et al.*, 2008).

after it ended. These include: 1.2 higher grades completed by women (but no effect for men); one-quarter of a standard deviation (SD) higher scores on reading comprehension tests for both men and women; and one-quarter SD higher scores on non-verbal cognitive ability tests for both men and women.[2] Two-stage least squares estimates that endogenise the actual supplement intakes (using exposure to the intervention as the identifying instrument) yield treatment-on-the-treated estimates that corroborate these magnitudes.

Section 1 describes the intervention and the areas in Guatemala where it was implemented. Section 2 provides a conceptual framework for modelling the effect of nutrition on educational attainment and Section 3 describes the data. The effects of the early childhood nutritional intervention on adult educational attainment are presented in Section 4. Section 5 provides conclusions.

1. The 1969–77 INCAP Nutritional Intervention

In the mid-1960s, protein deficiency was seen as the most important nutritional problem facing the poor in developing countries, and there was considerable concern that this deficiency affected children's ability to learn. The Institute of Nutrition of Central America and Panama (INCAP), based in Guatemala, was the locus of a series of studies on this subject, leading to a nutritional supplementation trial begun in 1969 (Habicht and Martorell, 1992; Read and Habicht, 1992; Martorell *et al.*, 1995*a*). The principal hypothesis underlying the trial was that improved pre-school nutrition would accelerate mental development. An examination of the effects on physical growth was included to verify that the nutritional intervention had biological potency (Martorell *et al.*, 1995*a*). To test the principal hypothesis, 300 rural communities with 500–1000 inhabitants in eastern Guatemala (in areas not directly affected by the civil war) were screened in an initial study to identify villages of appropriate compactness (so as to facilitate access to feeding centres – see below), ethnicity and language, diet, access to health care facilities, demographic characteristics, child nutritional status, and degree of physical isolation.

Using these criteria, two sets of village pairs (one pair of 'small' villages with about 500 residents each and another pair of 'large' villages with about 900 residents each)

[2] There have been previous studies of educational outcomes for this population, primarily focusing on earlier rounds of data collected (and thus only examining adolescents and young adults) or using select subsamples of adults. All of these studies employ a different methodological approach, in particular conditioning on grades completed without econometric controls (apart from also including directly other control variables) for its behavioural determinants. Because the intervention may have had direct effects on completed schooling (as we show for women), conditioning on completed schooling makes causal interpretation more problematic. As a result, direct comparison of our results with earlier ones is difficult. Pollitt *et al.* (1993) reports positive associations of the intervention on tests of general knowledge, numeracy reading comprehension, and vocabulary given in 1988–9, conditional on completed schooling levels. Li *et al.* (2003, 2004) examines measures of educational attainment among a selected subsample of adult women in the study villages who had borne a child between 1996 and 1999. They find that exposure to the intervention was associated with higher performance on a scale that combined literacy, numeracy and general knowledge, again conditional on completed schooling (in this case, completion of sixth grade). Stein *et al.* (2008) report results consistent with Li *et al.* (2003, 2004) for the separate test scores considered in this article, again conditional on completed schooling and without econometric controls for its behavioural determinants. Finally, in another study using the same survey and analytical design as ours, Hoddinott *et al.* (2008) examine the impact of exposure to the nutritional supplement on income earned per hour (wage), hours worked and total income, finding that exposure increased male wages by 40% but had no significant effect on the other outcomes.

were selected.[3] Before the intervention, the village pairs were similar in terms of a variety of nutritional, social and economic outcomes, though it turned out slightly less so in terms of educational outcomes. Child nutritional status before the intervention, as measured by length at three years of age, was similar across villages (Habicht *et al.*, 1995), and indicated substantial under-nutrition with over 50% severely stunted – height-for-age z-scores less than -3 (Martorell, 1992).[4] Maternal height was also not statistically different across villages (Rivera *et al.*, 1995). Specially collected village census data showed similar patterns of civil status of household heads, religious affiliation, agricultural employment and housing characteristics across the four villages. One village, however, had somewhat higher literacy and schooling levels for adults (Bergeron, 1992; Maluccio *et al.*, 2005*c*).

Two of the villages, one from within each pair matched on population size (i.e., one large, known as Conacaste, and one small, San Juan), were randomly assigned to receive as a dietary supplement a high protein-energy drink, *atole*. *Atole* comprised Incaparina (a vegetable protein mixture developed by INCAP and widely accepted for young children in Guatemala), dry skim milk and sugar, and had 163 kcal and 11.5 grams of protein per 180 ml cup. *Atole*, the Guatemalan name for porridge, was served hot and was slightly gritty but with a sweet taste.

In designing the intervention, there was considerable concern that the social stimulation for children – resulting from their social interactions while attending feeding centres where the supplement was to be distributed, the observation and measurement of their nutritional status, and the monitoring of their intakes of *atole* – also might affect child nutritional and cognitive outcomes, thus confounding efforts to isolate the nutritional effect of the *atole* supplement. To address this concern, in the two remaining villages, Santo Domingo (large) and Espíritu Santo (small), an alternative supplement, *fresco*, was provided, under identical conditions. *Fresco* was a fruit-flavoured drink, which was served cool and thus an appreciated refreshment in these areas, where average monthly temperatures ranged from 24 to 30 degrees Celsius. It contained no protein and only sufficient flavouring agents and sugar for palatability, and had about one-third of the calories of *atole* per unit volume (59 kcal/180 ml). Several micronutrients (iron, thiamine, riboflavin, niacin, ascorbic acid and vitamin A) also were added to both *atole* (which already had some) and *fresco*, in amounts that yielded equal concentrations across the supplements per unit of volume (Habicht and Martorell, 1992).[5]

The nutritional supplements (i.e., *atole* or *fresco*) were distributed in each village in centrally-located feeding centres and were available twice daily, to *all* members of the village on a voluntary basis, for two to three hours in the mid-morning and two to three

[3] There is little reason to believe the four villages ultimately selected were substantially atypical from the 300 potential candidates. For example, none of the four had had significant previous public health interventions (Habicht and Martorell, 1992). Information collected during the screening process on the other villages, however, is no longer available to explore this further.

[4] z-scores are used to normalise measured heights and weights against those found in well-nourished populations. They are age and sex specific; for example, a z-score of height-for-age is defined as measured height minus median height of the reference population, all divided by the standard deviation of the reference population for that age/sex category. Therefore a z-score of -3 means three standard deviations of the reference population below the reference median.

[5] For the first two years of the intervention, *atole* had a higher concentration of micronutrients. Given the short period over which micronutrient concentrations differed, however, it is not feasible to isolate the effect of those differences in the empirical analyses.

hours in the mid-afternoon, times selected to be convenient to mothers and children but that did not interfere with usual meal times. All residents of all villages also were offered high quality curative and preventative medical care free of charge throughout the intervention. Preventative services, including immunisation and anti-parasites campaigns, were conducted simultaneously in all villages.[6] To ensure that the results were not systematically influenced by the characteristics of the health, research, or survey teams, all personnel were rotated periodically throughout the four villages, each of which was separated by at least 10 kilometres.

From 1969 to 1977, INCAP implemented the nutritional supplementation and the medical care. While the supplement was freely available to *all* village residents (as described above), the associated observational data collection focused on children between zero and seven years of age at *any point during* the intervention period.[7] Thus all children under seven years of age residing in the villages at the start of the intervention, as well as those born in the villages during the intervention, were included in the survey, a total of 2,392 children. Data collected at the child level included precise measurement of actual daily supplement intakes (from which caloric and protein intakes can be calculated), periodic 24-hour food recall and periodic anthropometric measurements until the child reached seven years of age or until the survey data collection ended in 1977, whichever came first. Nevertheless, in cases where the child surpassed seven years of age first, he or she continued to be exposed to the intervention until it ended. Children in the sample, then, were all born between 1962 and 1977, and the type, timing and length of exposure for particular children depended on their village and date of birth.

For this study, where we use children's differential exposure to the availability of the nutritional supplements to identify the effect of *atole* relative to *fresco*, it is first important to establish that the two interventions resulted in differential consumption of calories, protein, and other nutrients. Approximately 70% of children between the ages of 0–36 months consumed at least some *atole*, with no difference between boys and girls. Similar overall participation rates were observed in *fresco* villages. Averaging over all children in the *atole* villages (regardless of their levels of voluntary participation), children 6–12 months consumed approximately 70 kcal of *atole* supplement per day; children 12–24 months, 90 kcal; and children 24–36 months, 120 kcal. Children in the *fresco* villages, however, consumed only 20 kcal of *fresco* supplement per day between the ages of 6–24 months, rising to approximately 30 kcal by the age of 36 months (Schroeder *et al.*, 1992).[8]

[6] For the interpretation and consideration of the external validity of our findings below, it is important to underscore the nature of the intervention, which involved intensive contact between researchers and villagers, as well as the provision of quality medical care. If these aspects of the intervention affect equally the impact of the two supplements, then the contrasts we explore below are externally valid to situations without the survey and medical care components of the intervention. If not, the observed effects may have been diminished or potentiated by these other aspects of the intervention (Habicht and Martorell, 1992).

[7] The intervention began in the larger villages in February 1969, and in the smaller villages, in May 1969. The nutritional supplements and medical care ended in all four villages at the same time, in February 1977, and the survey data collection ended seven months later (Martorell *et al.*, 1995*a*).

[8] Children less than four months of age in *atole* villages consumed very little supplement, and those who did were given a modified, age-appropriate mixture of skimmed milk and sugar. However, INCAP actively promoted the nutritional value of the supplements for pregnant and breastfeeding women, as well, so that young infants could benefit indirectly through the mother's improved nutritional status (Habicht and Martorell, 1992). For this reason, we include the period starting at birth as part of the exposure period.

To assess whether *total* intakes by these children increased, Islam and Hoddinott (2009) estimate an ordinary least squares (OLS) model in which the dependent variable is the sum of calories consumed at home (measured using periodic 24-hour recall dietary surveys) plus calories from supplements (measured directly at the feeding centres). In addition to a variety of controls for maternal and paternal characteristics and household characteristics similar to the ones we include in our analyses below, they include an indicator variable equal to one if the child resided in one of the two villages where *atole* was provided (and zero otherwise), to estimate the intent-to-treat effect of the *atole* intervention (relative to the *fresco* intervention) on intakes. For children aged 12–36 months they estimate total caloric consumption for children exposed to *atole* increased by approximately 10% and total protein intake by 40%. The intervention, therefore, increased energy and, in particular, protein intakes for young children in *atole* villages relative to *fresco* villages. In addition, for children under 36 months of age, the volume of *atole* consumed was higher than the volume of *fresco* consumed, with the implication that intakes of micronutrients were also greater for children in *atole* villages. The *atole* intervention succeeded in improving nutrient intakes in general, rather than just protein as originally was envisioned in 1969.[9]

2. Conceptual Framework and Identification Strategy

We treat investments in nutrition, health, and education as part of a life cycle dynamic programming problem solved by the family of the individual, subject to the constraints imposed by parental family resources and options available in the community to the individual as he or she ages; see Cunha *et al.* (2005) for a formal statement. The family's optimisation problem can be solved to obtain a relation that we interpret as a reduced-form equation for the determinants of each education-related outcome (E_i) for each individual i:

$$E_i = f(N_i, N_i^A, \mathbf{I}_i, \mathbf{F}_i, \mathbf{C}_i^f, \mathbf{C}_i^v, \varepsilon_i). \qquad (1)$$

For our analyses, outcome variables were measured in itwo follow-up rounds (in 1988–9 and 2002–4) while most explanatory variables are based on the initial 1969–77 survey (or are retrospective) and can be considered lagged values. We employ a

[9] A possible criticism of interpreting our results as the effect of nutrition is that the nutritional supplements may have had large income effects. This is unlikely, however, for several reasons. First, the behaviour of villagers did not suggest that the supplements were of significant monetary value. Despite the fact that supplements were freely available every day to all inhabitants of the communities, few men or school-age children frequented the feeding centres, even on weekends when the opportunity cost of their time in terms of work or school presumably was lower. Likely because INCAP actively promoted the nutritional value of the supplements for pregnant and breastfeeding women and young children, however, attendance among these groups was higher. Second, the actual monetary value of the supplements was low. We estimate the cost of the ingredients for one cup of *atole* and one cup of *fresco* to have been US$ 0.018 and 0.004. Mean household incomes were approximately US$ 400 in 1975 (Bergeron, 1992). Thus one year's worth of a daily cup of *atole* (US $6.60) and of *fresco* (US $ 1.50) was approximately 1.7% and 0.4% of average annual household income, and on average children 0–36 months of age consumed less than this. The literature on the impact of cash transfers on human capital suggests that these amounts are not large enough to impact significantly on child health and education (Haddad *et al.*, 2003). It is likely that the medical care, however, had a greater income effect for households, but this effect was equally present in both *atole* and *fresco* villages. Lastly, there is direct biological evidence that the *atole* had nutritional effects. Biochemical analyses of blood showed that biomarkers of nutrient status were improved following ingestion of *atole* (Habicht *et al.*, 1973).

double-difference estimation strategy to estimate the intent-to-treat effect of improved nutrition by contrasting individuals exposed at different periods of their lives in *atole* versus *fresco* villages, as described below.

Specifically, we first construct for each individual a dummy indicator variable reflecting whether he or she was exposed to *either* intervention for the entire period from birth to 36 months of age, based on their birth date and the dates of operation of the interventions in each village. This indicator variable (N_i) represents a main 'cohort' effect that captures factors common to all children in any of the villages in this age range, including improved medical services and increased social stimulation present under both the *atole* and the *fresco* interventions. Next, we interact this main effect with an indicator variable of whether or not the individual lived in one of the two *atole* villages. This second indicator (N_i^A), exposed to *atole* from birth to 36 months, captures the differential effect of being in an *atole* village in comparison with those in the same cohort in a *fresco* village.

There are also individuals who were only partially exposed during the period from birth to 36 months, to either *fresco* or *atole*. These include persons born before the intervention began (for example, those born in 1967 or 1968) and others, born after 1974, exposed from birth until the end of the intervention but not for their entire first 36 months. To the extent *atole* (relative to *fresco*) also has an effect on individuals exposed partially from birth to 36 months (or at ages older than 36 months), our estimates are likely to underestimate the full programme effect, a potential bias we examine in Section 4.4. Because for the basic analyses we use potential exposure, which is not conditional on actual participation or intakes, our estimates yield the intent-to-treat effect of exposure to the high-protein, high-energy supplement, *atole*, versus the no-protein, low-energy supplement, *fresco*. Were we to include directly an indicator variable for treatment type (e.g., for *atole*) and there were no other covariates, the coefficient on N_i^A would be the double-difference estimator where the first difference would be the difference in outcomes for those exposed to *atole* during their entire first 36 months and those exposed to *fresco* during their entire first 36 months, and the second difference would be those exposed to *atole* during other periods of their life and those exposed to *fresco* during other periods of their life. Rather than include an indicator variable for *atole*, however, we include indicators for three of the four villages ($\mathbf{C}_i^f$), capturing all fixed characteristics of these villages that might affect education-related outcomes. It is crucial to include these because of the small number of villages in the experiment. They control for factors such as the initial sizes of the villages, persistent cultural differences and fixed differences in educational or economic opportunities that might result in different educational investments across villages, even in the absence of the interventions.

We chose for our basic specification an exposure period from birth to 36 months of age based on earlier research with the 1969–77 data (we describe results using alternative exposure periods in Section 4.5). A key finding has been that growth failure occurred primarily in utero and in the first two to three years of life, and was the cause of the short stature of adults. The gap between the Guatemalan sample and the international reference population increased until about three years of age, but remained fairly constant thereafter (Martorell *et al.*, 1995*b*). Supplementation with *atole*, in comparison to *fresco*, increased the height of three-year-old children by about

2.5 centimetres and reduced by half the prevalence of severe stunting. Supplementation produced its biggest effects by two years of age and after three years of age did not influence child growth rates (Schroeder *et al.*, 1995).[10] A further justification for using the 0–36 month exposure period is that it is viewed as a crucial window for interventions to stimulate positive cognitive development, particularly in settings where undernutrition and poor health are common (Grantham-McGregor *et al.*, 2007). A final justification for using the 0–36 month exposure is that this period also was found to be relevant in assessing the effect of *atole* on wages earned by men in the sample, using a similar identification strategy (Hoddinott *et al.*, 2008).

Identification of the programme effect, i.e., the effect of switching from *fresco* to *atole*, then, rests in part on randomisation of villages into one of the two interventions (though it is important not to overemphasise this aspect since there are only two villages in each group), as well as on the comparison across age-cohorts who were exposed to those interventions at different ages. While this latter aspect cannot be regarded as random, we argue it is exogenous to household behaviour in that the start and end dates of the intervention were externally determined (and for the end date, somewhat of an abrupt 'surprise' when National Institutes of Health (NIH) funding was discontinued). We also examined whether the *atole* intervention had an effect on fertility, by estimating a double-difference model comparing the fertility of younger women in *atole* versus *fresco* villages with older women (who had completed their fertility by the time the intervention began) in *atole* versus *fresco* villages, and found no effect. Our maintained assumptions for identification, therefore, are that exposure to the interventions during other periods of life has little (or equal) impact on the outcomes of interest and that there are no 'random' age-cohort-village effects beyond the controls we include (Section 3.3) that affect the educational outcomes of interest. We carry out a series of alternative identification strategies that provide evidence that the latter component of the maintained assumption is valid (Section 4.4).

The other determinants in the model include: a vector of individual characteristics such as indicators of sex and birth year ($\mathbf{I}_i$), a vector of fixed family background characteristics ($\mathbf{F}_i$), a vector of varying community-level characteristics related to schooling availability and quality at different ages ($\mathbf{C}_i^v$), and a disturbance term (ε_i) that affects the educational outcome of interest. The disturbance term includes all unobserved or unmeasured variables. These may include some fixed individual and family variables that affect the education-related indicators of interest directly and that are correlated with the included observed family background variables; for some related evidence, see Behrman *et al.* (2008). As a result, we do not emphasise the coefficient estimates of family background variables in the discussion but rather provide evidence that they are controls for variation unrelated to the intervention (Section 4.4) which allow more precise estimation. The control variables are described in more detail in Section 3.3.

[10] The nutritional literature emphasises that undernutrition is most common and severe during periods of greatest vulnerability (Martorell, 1997; UNICEF, 1998). One such period is the first two to three years of life. Young children have high nutritional requirements, in part because they are growing so fast. The diets commonly offered to young children in developing countries to complement breast milk are monotonous and have low energy and nutrient density; as a result, multiple nutrient deficiencies are common. Moreover, young children are susceptible to infections because their immune systems (which are both developmentally immature and potentially compromised by poor nutrition) fail to protect them adequately. Foods and liquids are often contaminated and lead to infections, which both reduce appetites and increase metabolic demands.

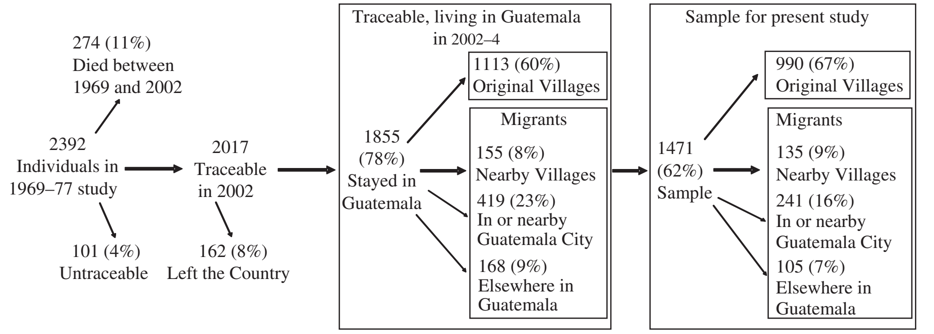

Fig. 1. *Relation Between Original 1969–77 Sample and the HCS 2002–4 Sample*
Updated from Grajeda *et al.* (2005, Figure 1).

3. Data

3.1. *The Follow-up (1988–9) and Human Capital Studies (2002–04)*

Subsamples of the original 2,392 children surveyed during 1969–77 have been re-surveyed periodically, including two surveys used in this article:

(1) the Follow-up Study in 1988–9 (Martorell *et al.*, 1995*a*); and
(2) the Human Capital Study (HCS) in 2002–4 (Grajeda *et al.*, 2005; Martorell *et al.*, 2005).

Our primary analyses use HCS because, by then, all individuals from the original 1969–77 survey were adults from 25–42 years old but in some analyses we use the Follow-up (1988–9) data, in part to assess potential attrition biases. Figure 1 shows what happened to the 2,392 individuals in the original 1969–77 sample by the time of the 2002–4 HCS: 1,855 (78%) were determined to be alive and known to be living in Guatemala: 11% had died – the majority due to infectious diseases in early childhood; 7% had migrated abroad; and 4% were not traceable (Grajeda *et al.*, 2005). The basic analyses using HCS include 1,471 respondents (53% female) who completed the questionnaire module pertaining to schooling.[11] They comprise 79% of the 1,855 individuals who were known to be alive and living in Guatemala, 73% of those known to be alive, and 62% of the original sample of 2,392. Measured from 1977 to 2002, even the most conservative measure (62%), which indicates an annual attrition rate of approximately 2%, is low when compared to shorter-term longitudinal surveys in developing countries (Alderman *et al.*, 2001*a*) or to long-term longitudinal surveys in the US (Fitzgerald *et al.*, 1998*b*).[12] Nevertheless, almost 40% represents substantial attrition, so we carefully assess potential biases due to attrition in Section 4.6.

Compared with a nationally representative sample of Guatemalans, individuals re-interviewed in HCS were relatively well off. This is unsurprising, however, since the

[11] Twenty-three (1.6%) of those completing the schooling module did not complete the reading comprehension or non-verbal cognitive ability modules (see Section 3.2), so that estimates for those dependent variables are based on 1,448 observations.

[12] Typically measures of attrition refer to households or individuals who were past infancy and early childhood when the sample was taken, so they would not even include the effects of infant and early childhood mortality that account for over a quarter of the attrition in the data used for this study.

study population is *ladino* and speaks Spanish as their first language, and poverty in Guatemala is concentrated among the large indigenous population (World Bank, 2003). On average, sample members had an expenditure-based poverty rate of 35% and an extreme poverty rate of only 3%, against national averages of 56% and 15%, respectively (Maluccio *et al.*, 2005*a*). Basic descriptive statistics are shown in Table 1. In general, the outcome variables are taken from the follow-up surveys while the explanatory variables are drawn from individual and household level data collected from 1969 to 1977, supplemented by community-level information collected both during earlier studies and retrospectively in 2002.

3.2. *Dependent Variables: Educational Outcomes*

We examine three aspects of education (E_i) across the life cycle:

(1) attained (or completed) schooling;
(2) reading comprehension; and
(3) non-verbal cognitive ability.

In this section, we focus on measures from HCS (2002–4), but those from the 1988–9 survey are included in Table 1 for later reference.

Attained schooling is measured as the highest grade completed. The formal educational system in Guatemala is divided into primary, secondary and post-secondary education. Primary school comprises grades one to six and children are expected to enrol in the calendar year in which they turn seven years old. Secondary school consists of five to seven grades, divided into two parts. The first three grades of lower secondary school (to grade nine) are the so-called 'basic' grades, and instruction is expected to provide academic and technical skills necessary to join the labour force. The fourth to seventh grades of upper secondary school are the so-called 'diversified' grades, and students can choose from among four tracks:

(1) general academic high school education, known as *bachillerato*;
(2) primary school teaching;
(3) technical education, such as a secretarial degree; and
(4) commercial education, such as an accounting degree.

Typically, students who plan to continue to university finish their *bachillerato* in two years at the secondary diversified level, thus completing 11 grades before going on to university. The other specialised degrees at the diversified level usually take three years (World Bank, 2003).

Over 95% of the HCS respondents started school, with no difference between males and females. Conditional on starting school, approximately 9% per year dropped out at the end of grades one through five and 30% stopped attending after completing the full six grades of primary education (Figure 2). As a result, less than 20% continued on to secondary school. Finally, only a small number of individuals in the sample (less than 3% of the sample) continued beyond secondary school to attend university or do advanced studies in technical fields. Apart from formal schooling, it was also possible to complete (primary and secondary school) grades via informal schooling, in particular adult literacy programmes. Our overall measure of grades completed includes both

Table 1

Means and Standard Deviations, By Sex

Variable	Full sample	Women	Men
Ever attended (formal) school	0.955	0.950	0.959
Age when started school (if ever attended) [$n = 1{,}365$]	6.799	6.782	6.815
	(1.09)	(1.00)	(1.19)
Completed first grade	0.861	0.854	0.869
Completed sixth grade	0.460	0.386	0.546***
Completed ninth grade	0.132	0.112	0.155**
Highest grade completed (measured in HCS)	4.698	4.303	5.150***
	(3.45)	(3.31)	(3.56)
Highest grade completed (measured in 1988–9) [$n = 1{,}323$]	3.884	3.689	4.075***
	(1.91)	(1.86)	(1.93)
Reading comprehension (SIA) in HCS [$n = 1{,}448$]	36.051	34.407	37.945***
	(22.31)	(21.85)	(22.71)
Reading comprehension (SIA) in 1988–9 [$n = 1{,}368$]	32.065	31.618	32.511
	(21.52)	(21.46)	(21.58)
Non-verbal cognitive ability (Raven *et al.*, 1984) in HCS [$n = 1{,}448$]	17.703	16.263	19.361***
	(6.14)	(5.40)	(6.51)
Non-verbal cognitive ability (Raven *et al.*, 1984) in 1988–9 [$n = 1{,}368$]	11.197	10.340	12.051***
	(4.59)	(4.19)	(4.81)
Calories from supplement 0–36 months (kcal/day)	36.542	34.125	39.314
	(61.93)	(59.55)	(64.48)
Protein from supplement 0–36 months (grams/day)	2.235	2.111	2.377
	(4.48)	(4.30)	(4.67)
Exposed to intervention			
From birth to 36 months (N_i)	0.394	0.372	0.419*
From birth to 36 months × *atole* (N_i^A)	0.211	0.206	0.218
Lived in San Juan (small *atole* village)	0.228	0.237	0.218
Lived in Conacaste (large *atole* village)	0.304	0.298	0.311
Lived in Espíritu Santo (small *fresco* village)	0.210	0.213	0.207
Lived in Santo Domingo (large *fresco* village)	0.258	0.253	0.264
Birth year	1970.1	1969.9	1970.2
	(4.21)	(4.30)	(4.11)
Male	0.466	–	–
Mother's age when child was born	27.487	27.646	27.305***
	(7.07)	(7.18)	(6.94)
Father's age when child was born	32.856	32.963	32.733
	(8.25)	(8.44)	(8.04)
Mother's highest grade completed	1.349	1.243	1.472
	(1.67)	(1.59)	(1.75)
Father's highest grade completed	1.679	1.663	1.700
	(2.09)	(1.99)	(2.20)
Household wealth index score	−3.094	−3.106	−3.080
	(0.89)	(0.88)	(0.91)
Village school permanent structure when child was 7	0.477	0.464	0.491
Village school permanent structure when child was 13	0.825	0.809	0.842*
Village student-teacher ratio when child was 7	39.935	40.277	39.543
	(8.96)	(9.35)	(8.48)
Village student-teacher ratio when child was 13	36.307	36.271	36.347
	(5.26)	(5.34)	(5.17)
N (unless otherwise indicated)	1,471	786	685

Notes. Except where indicated, sample consists of 1,471 individuals who were exposed to the INCAP supplementation intervention between 1969 and 1977 and who were subsequently re-interviewed in HCS 2002–4. Standard deviations reported in parentheses. * indicates difference between men and women is statistically significant at a 10% level; ** significant at the 5% level; *** significant at 1% level.

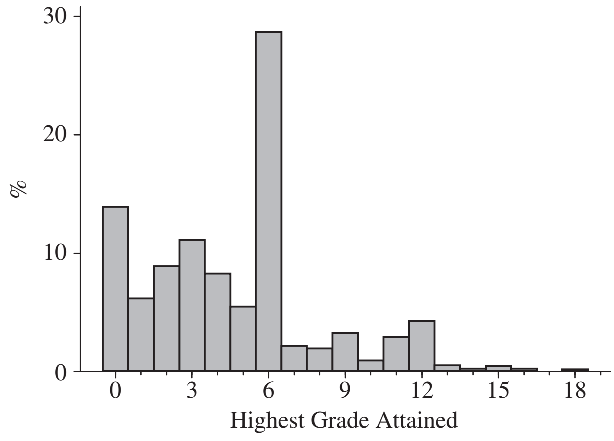

Fig. 2. *Distribution of Highest Grade Completed*

types of schooling, though informal schooling is relatively uncommon for this population, with only 15% of the respondents reporting having ever attended. Other salient characteristics of schooling attainment in the sample are that:

(1) men complete an average of 0.9 grades more than women (Table 1);
(2) schooling attainment has increased over time for both men and women; and
(3) comparisons of unconditional means always show that grade attainments are higher in Espíritu Santo, the small *fresco* village, than in the other villages, particularly for men (Stein *et al.*, 2005). This partly reflects the fact that Espíritu Santo has long benefited from its proximity to the municipal capital town of El Jícaro, where primary and secondary schools were present even before 1962, the birth year of the oldest individuals in the sample. Indeed, parental schooling is also higher in Espíritu Santo than in the other villages (Maluccio *et al.*, 2005*c*).

Reading comprehension was measured via a two-part standardised test. Respondents who reported having passed fewer than four grades of schooling, or those who reported four to six grades of schooling but could not correctly read aloud the headline of a local newspaper article, were first given a literacy test. Individuals who passed this literacy screen, or who reported more than six grades of schooling (and thus were presumed to be literate) then took the Inter-American Series test vocabulary and reading comprehension modules (*Serie Interamericana* or SIA, for its acronym in Spanish). The SIA test was designed to assess reading skills of Spanish-speaking children in Texas (Manuel, 1967) and includes several levels of difficulty. Based on pilot testing for HCS, Level 2 for comprehension and Level 3 for vocabulary were used (approximately 3rd and 4th grade equivalents). The same tests (but Level 2 for vocabulary) were implemented in 1988–9. In the earlier survey, the tests demonstrated adequate test-retest reliability with correlation coefficients above 0.85 (Pollitt *et al.*, 1993). The reading comprehension module had 40 questions and the vocabulary module 45 questions, yielding a maximum possible score of 85 points. Questions on the test become progressively harder. Those who did not pass the literacy screen

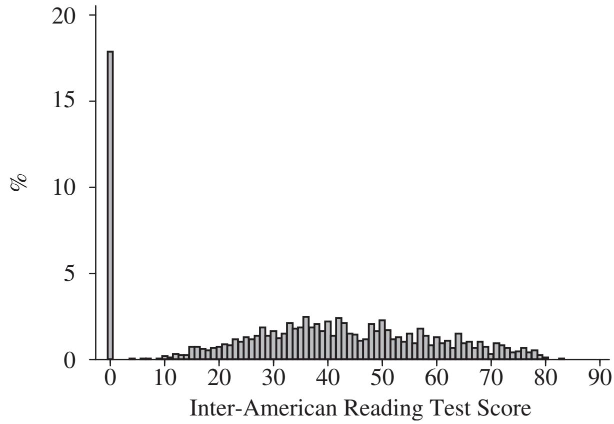

Fig. 3. *Distribution of Reading Comprehension (SIA) Test Scores*

pre-test were given a zero (18% of the sample) and all others the number of correct answers. Figure 3 shows the distribution of the SIA test scores for the HCS sample.

All individuals (regardless of results on the literacy screen) were administered the Raven's Progressive Matrices test (hereafter Raven's test), an assessment of non-verbal cognitive ability (Raven *et al.*, 1984). Raven's tests are considered to be a measure of eductive ability – 'the ability to make sense and meaning out of complex or confusing data; the ability to perceive new patterns and relationships, and to forge (largely non-verbal) constructs which make it easy to handle complexity' (Harcourt Assessment, 2008). The test consists of a series of pattern-matching exercises with the respondent asked to supply a 'missing piece' and with the patterns getting progressively more complex. We administered the first three of five scales (*A*, *B* and *C* with 12 questions each for a maximum possible score of 36), since pilot data suggested, and subsequent survey data confirmed, that few respondents were able to progress beyond the third scale. These same components of the Raven's test also were implemented in 1988–9 and demonstrated adequate test-retest reliability, with a correlation coefficient of 0.87 (Pollitt *et al.*, 1993). Respondents could take as long as they needed, with a typical test completed in approximately 40 minutes. Raven's tests were scored as the number of correct answers summed across the three scales and the mean score in HCS was 17.7 with a SD of 6.1 (Table 1).

In all the regression analyses reported below, reading comprehension test (SIA) scores and non-verbal cognitive ability test (Raven's) scores are expressed as z-scores standardised to have mean 0 and SD 1 within the sample.

3.3. *Control Variables*

In addition to the village fixed effects, the cohort control indicating whether the individual was exposed to either intervention from birth to 36 months of age (N_i) and the measure of exposure to the *atole* nutritional intervention for that same period (N_i^A) (all described in Section 2), we also control for individual characteristics ($\mathbf{I}_i$), a vector of fixed family background characteristics ($\mathbf{F}_i$), and varying community-level characteristics ($\mathbf{C}_i^v$).

We include two individual-level characteristics, sex and birth year. The latter is important as it captures secular trends common to all individuals that might affect educational outcomes, beyond the cohort control. We also consider a variety of specifications in which we allow the effect of year of birth to be more flexible (Section 4.4). The vector of fixed family background characteristics include mother's and father's ages when the child was born, mother's and father's completed grades of schooling, an index of household wealth at the time of the intervention and indicator variables capturing the distance from the household to the feeding centre where the supplements were distributed and consumed. As part of the survey work accompanying the intervention in 1969–77, all households in these villages – including those with children participating in the supplementation trial – were included in village censuses in 1967–8 and again in 1975 that ascertained ownership of a set of household durables as well as housing characteristics. Using principal components, these assets and characteristics were combined into an index we interpret as a 'wealth' measure. Though this index surely misses some dimensions of wealth such as financial and productive assets, at the time in the late 1960s and early 1970s in these villages, such non-measured assets were not only uncommon but also likely to be highly correlated with the assets that were measured (Maluccio *et al.*, 2005*b*). Indicators of proximity to the feeding centre in each village account for differences in accessibility that have been shown to be associated with supplement intakes (Schroeder *et al.*, 1992) even though, by design, fewer than 20% of the villagers lived more than one kilometre from the centrally located feeding centre.

Finally, we exploit detailed qualitative and archival data found in Pivaral (1972) and Bergeron (1992), as well as a specially commissioned retrospective study undertaken in 2002 (Estudio 1360, 2002), to control for some important community-level characteristics that have changed over time ($\boldsymbol{C}_i^v$). Specifically, we construct community-level covariates that relate as closely as possible to the timing of key education-related decisions in a child's development. In the estimates reported below, we include four such controls: indicator variables of the availability in the village of a permanent (cement-block) structure for the primary school when the respondent was seven and 13 years old, and primary school student-teacher ratios when the respondent was seven and 13 years old. We use age seven because that is the age at which most children were starting primary school and age 13 because that is the age at which children would have begun secondary school if they had entered at age seven and progressed one grade each year through primary school. While these variables reflect community-level characteristics, they vary by single-year age cohorts within each village. This is an improvement over the more typical approach of including measures of such variables at a given time for a population with different ages at that point in time, since it more closely relates the availability of schools to the periods in the child's life when critical decisions about schooling were being made.[13] Unless indicated otherwise, the same set of controls is used in all regression results reported.

[13] We also considered a variety of other community characteristics as controls, including exogenous income shocks to these communities (booms and busts in the prices of key agricultural commodities; the availability of new sources of wage employment), changes in infrastructure (roads, water and electricity) and other dimensions of school quality (e.g., damage suffered by schools as a result of a severe earthquake in 1976). As their inclusion did not alter substantially the results reported here, they are not included in the reported specifications.

3.4. *Calculation of the Standard Errors*

The original intervention occurred in only four villages and duration of exposure to the intervention was dependent on date and village of birth. All regression results include both village fixed effects and a covariate capturing cohort effects, which also capture all fixed components of possible intra-village and intra-cohort correlations. Moulton (1990, p. 334), however, has noted that '[i]t is reasonable to expect that units sharing an observable characteristic, such as industry or location, also share unobservable characteristics that would lead the regression disturbances to be correlated'. These correlations, if positive (and in our case, also varying over time because we control for village fixed effects), may cause the estimated standard errors to be biased downward. Throughout the article, standard errors are calculated allowing for clustering at the mother level (there are 597 distinct mothers). For the principal findings reported below, however, we considered several alternative ways of calculating the standard errors to address the implications of possible serial correlation (e.g., time varying correlation among the errors within villages or autocorrelation of errors over time within villages).

First, we considered clustering at the village birth-year level (yielding 64 clusters = 4 villages $\times$ 16 different birth years). Work by Angrist and Lavy (2002) and Wooldridge (2003), however, suggests that standard corrections for clustering are valid only when the number of groups or clusters is large and we cannot be sure that $N = 64$ is large enough.[14] In light of this, following Bertrand *et al.* (2004), we block bootstrapped the standard errors, using the same 64 clusters. We also implemented the procedure proposed by Donald and Lang (2007) for differences-in-differences with a small number of groups, again treating village birth-years as different groups. This approach, a two-step estimator based on the village birth-year means (after conditioning out variables that vary at the individual level) involves a considerable loss in degrees of freedom as the estimating sample size drops from approximately 1,450 to 64 observations. Finally, we implemented a Newey-West correction procedure (Statacorp, 2007), aggregating observations (within each village) for each quarter (of each year). All the principal findings remained statistically significant (at 5%) across the various approaches. Moreover, calculating standard errors based on clustering at the mother level turned out to be the most conservative approach. For this reason, we report those results. We note, however, that the standard errors are likely to be too small if there is substantial serial correlation across cohorts, since none of the methods we employ completely address this problem. On the other hand, the fact that we find similar results across different outcome measures raises the confidence that the results are not simply driven by sampling variation.

4. The Effect of Childhood Nutrition on Adult Educational Outcomes

4.1. *Schooling Attainment*

Table 2 presents the base specification we use to explore the intent-to-treat effect of the early childhood nutritional intervention a quarter century later on highest grade

[14] For this reason, we do not cluster at the village level with number of clusters equal to four, which led to highly significant results for all the principal findings.

Table 2

OLS Estimates of Impact of Exposure to Atole (Relative to Fresco) from Birth to 36 Months on Highest Grade Completed Using HCS (2002–4), By Sex

Covariates	Full sample	Women	Men
Exposed to intervention			
From birth to 36 months	−0.009	−0.214	0.182
	(0.228)	(0.362)	(0.345)
From birth to 36 months × *atole*	0.349	1.168**	−0.432
	(0.362)	(0.548)	(0.534)
Controls			
Lived in San Juan (small *atole* village)	0.137	−0.259	0.374
	(0.438)	(0.526)	(0.641)
Lived in Conacaste (large *atole* village)	−0.439	−0.617	−0.346
	(0.370)	(0.439)	(0.554)
Lived in Espíritu Santo (small *fresco* village)	1.731***	1.413***	1.976***
	(0.382)	(0.481)	(0.500)
Birth year	0.011	−0.001	0.041
	(0.037)	(0.045)	(0.055)
Male	0.710***	–	
	(0.171)		
Mother's age when child was born	−0.014	−0.018	−0.011
	(0.020)	(0.023)	(0.027)
Father's age when child was born	0.014	0.003	0.027
	(0.016)	(0.018)	(0.021)
Mother's highest grade completed	0.392***	0.407***	0.391***
	(0.063)	(0.095)	(0.083)
Father's highest grade completed	0.281***	0.216***	0.335***
	(0.061)	(0.082)	(0.068)
Household wealth index score	0.852***	0.717***	0.970***
	(0.147)	(0.204)	(0.163)
Village school permanent structure when child was 7	0.070	0.081	0.079
	(0.284)	(0.387)	(0.395)
Village school permanent structure when child was 13	0.043	0.010	−0.080
	(0.311)	(0.362)	(0.523)
Village student-teacher ratio when child was 7	−0.025**	−0.036**	−0.009
	(0.012)	(0.015)	(0.019)
Village student-teacher ratio when child was 13	−0.006	−0.010	−0.001
	(0.019)	(0.026)	(0.032)
Constant	−14.376	11.82	−73.80
	(72.900)	(88.392)	(107.733)
R^2	0.243	0.221	0.276
N	1,471	786	685

Notes. Sample consists of all individuals who were exposed to the INCAP supplementation intervention between 1969 and 1977 and who were subsequently re-interviewed in HCS 2002–4 (with valid completed grades of education data). Additional covariates included but not reported are indicator variables for distance to feeding centres and for observations with missing data on each of the following: maternal age, paternal age, maternal schooling, paternal schooling, household wealth index, and distance to feeding centre. Standard errors are calculated allowing for clustering at the mother level and reported in parentheses (StataCorp, 2007). * Significant at the 10% level; ** significant at the 5% level; *** significant at 1% level.

completed for adults interviewed in HCS (2002–4) when they were between the ages of 25 and 42. We use an OLS estimator and calculate standard errors that allow for clustering at the mother level (Section 3.4). Estimates for women and men combined are shown in the first column, and for women and men separately in the second and third columns. Reflecting the discussion above, conditional on the other variables, individuals born in Espíritu Santo had significantly higher attained schooling, underscoring the importance of including village fixed effects. Consistent with patterns seen in Guatemalan society as a whole, men completed more schooling on average than women. Parental schooling and resources as measured by the wealth index also have positive significant associations with the completed grades for the child, though these may be biased upward if the scores partly reflect intergenerationally correlated 'ability' endowments also associated with the parental schooling or wealth (Behrman *et al.*, 2008). Although formal primary education has been available in all of the villages since the early 1960s, since then the quality of that schooling has improved differentially across villages (Maluccio *et al.*, 2005*c*). The presence of a permanent schooling structure in the village does not seem to have had an influence on grades completed in the sample. The student–teacher ratio, however, does. This effect is apparent at younger ages and appears to have been more important for women than for men.

For the combined sample, we do not find a significant intent-to-treat effect of the *atole* intervention (relative to the *fresco* intervention) for children exposed from birth to 36 months of age. A Chow test weakly indicates that the models are best estimated together rather than separately by sex ($p = 0.13$); however, one of the two statistically significantly different coefficients for women versus men in a fully interacted (with sex) model is for the *atole* exposure variable. When split by sex, there is a large and significant effect of *atole* for women – more than one full grade, or a third of a SD in the sample. This larger impact for girls is consistent with evidence that in many developing countries girls' schooling is more responsive to changes in determinants than boys' schooling (World Bank, 2001).

To understand what underlies these results better, as well as to verify that the findings are not an artefact of using OLS in a situation where the dependent variable has a non-normal distribution (Figure 2), we explored the point in the schooling cycle at which the intervention had an effect.[15] In results not shown, we estimated separate probits that considered six distinct outcomes: ever attended school; completed grade three (or more); completed grade six (or more), i.e., completed primary school; completed grade seven (or more), i.e., entered secondary school; completed grade nine (or more), i.e., completed lower secondary school; and completed grade 11 (or more), i.e., completed high school (for those following the *bachillerato* course). Women exposed to the *atole* supplement were more likely to have completed primary school ($p = 0.10$), entered secondary school ($p = 0.11$) and completed lower secondary school ($p = 0.09$). For men, however, there were no significant patterns. The intervention had the effect of promoting greater equality of schooling attainment between women and men, compared to their parents, as it was more effective in the subpopulation (i.e., women) with lower previous schooling.

4.2. *Reading Comprehension*

In the top panel of Table 3, we present the intent-to-treat effect of the intervention on reading comprehension (as measured by z-scores of the SIA test), which are likely to be strongly influenced by schooling. They show for the full sample that exposure to the

[15] We also examined whether there was an effect on age at starting school for either males or females, but found none (results not shown).

Table 3

Estimates of Impact of Exposure to Atole (Relative to Fresco) from Birth to 36 Months on Reading Comprehension (SIA) and Non-verbal Cognitive Ability (Raven's) Using HCS (2002–4), By Sex

	Full sample	Women	Men
Reading comprehension (SIA)			
SIA test z-score (OLS)	0.277**	0.394**	0.212
	(0.110)	(0.158)	(0.163)
Probit (marginal effect):			
(1) if second quartile or higher	0.096*	0.142*	0.068
	(1.87)	(1.91)	(0.93)
(1) if third quartile (above median) or higher	0.140**	0.130	0.184**
	(2.47)	(1.63)	(2.17)
(1) if top quartile	0.115**	0.190**	0.053
	(2.19)	(2.31)	(0.72)
Non-verbal cognitive ability (Raven's)			
Raven's test z-score (OLS)	0.238**	0.210	0.272
	(0.118)	(0.156)	(0.177)
Probit (marginal effect):			
(1) if second quartile or higher	0.039	0.032	0.047
	(0.82)	(0.45)	(0.73)
(1) if third quartile (above median) or higher	0.100*	0.128*	0.078
	(1.94)	(1.87)	(0.97)
(1) if top quartile	0.173**	0.153**	0.183**
	(3.25)	(2.09)	(2.45)
N	1,448	775	673

Notes. Sample consists of all individuals who were exposed to the INCAP supplementation intervention between 1969 and 1977 and who were subsequently re-interviewed in 2002–4 (with valid scores on both SIA and Raven's tests). Additional covariates included but not reported are the same as those in Table 2. Standard errors are calculated allowing for clustering at the mother level and reported in parentheses (Statacorp, 2007). For probits, we present the average marginal effect (and z-value) calculated using methodology proposed by Norton *et al.* (2004). * Significant at the 10% level; ** significant at the 5% level; *** significant at 1% level.

atole intervention during childhood significantly and substantially increased reading comprehension as an adult. Using the point estimate on the linear pooled specification, the intervention led to an average improvement of 0.28 SD, over one-quarter of a SD. The estimated effects are nearly twice as large for women as for men, and estimated effects for men do not appear to be significant (and a Chow test suggests separate estimation is appropriate, with $p < 0.01$).

These OLS estimates, however, neglect both the censoring of the SIA test scores at zero and the fact that those who did not pass the literacy screen were given a zero on the test (before converting into z-scores), as well as the fact that questions on the test become increasingly difficult. For these reasons, we provide the results of three additional specifications, respecifying the outcome variable by dividing the sample into quartiles based on the SIA test score. Marginal effects for probit results for those scoring in the second quartile or above, third quartile or above, and fourth (and highest) quartile are shown. For the combined sample, exposure to *atole* led to increased probability of scoring in a higher quartile on the SIA test for all three specifications. The effects appear to be larger and more precisely estimated for women but men who were exposed were also 18 percentage points more likely to be in the top half of the distribution. Even though the intervention appears not to have increased grades completed for men, it did increase the probability that they scored above the median on the SIA test. This may be the result of greater learning during the same number of grades, though it also could reflect learning during post-schooling experiences, for example in the labour market (Behrman *et al.*, 2008).

4.3. *Non-verbal Cognitive Ability*

In the bottom panel of Table 3, we present the impact of the intervention on z-scores of the Raven's test of non-verbal cognitive ability. We find that for the combined sample, exposure to *atole* had a significant effect on non-verbal cognitive ability measured in adulthood. The effects for women and men, considered separately, are similar in magnitude, and tests suggest that in this case pooling the sample is preferred (the p-value on the Chow test is 0.38). The findings also hold when we consider different formulations of the Raven's test score that take into account that the test gets progressively more difficult (i.e., probits on whether the score was in higher quartiles as done for the SIA test). For women and men separately, there are significant increases of 15 percentage points or more in the probability of scoring in the top quartile.

The average intent-to-treat effect of the *atole* intervention, 0.24 SD, represents nearly one quarter of a SD, similar in magnitude to the results for reading comprehension. These results (as well as the general increase in Raven's test scores when compared to the 1988–9 survey shown in Table 1) raise questions about the often made interpretation of tests like these as reflecting innate abilities that cannot be altered by household and community allocations of resources (Alderman *et al.*, 1996; Boissiere *et al.*, 1985).

4.4. *Alternative Identification Strategies*

Our identification strategy relies heavily on observational differences across cohorts, and the assumption that we control for all relevant cohort and intervention (*atole* or *fresco*) level factors. In this Section, using a variety of different identification strategies that alter these assumptions (but include the same set of control variables unless otherwise indicated), we consider the robustness of the results for HCS reported in Tables 2 and 3 for the three significant outcomes:

(1) schooling attainment for women;
(2) SIA test z-scores for combined men and women; and
(3) Raven's test z-scores for combined men and women.

(We also report, but do not discuss in detail, the always insignificant results for schooling attainment for men.) For all of these outcomes, the estimated main 'cohort' effect for those exposed (i.e., the coefficient on N_i) is always insignificant and relatively small (less than a third) compared to the effect of *atole* (i.e., the coefficient on N_i^A). In addition, the estimated effect on the birth year trend is also small and always insignificant. These patterns suggest there are no important unobserved cohort effects potentially biasing the results. As a further check on whether the results are changed with more stringent controls for cohort effects, we re-estimate the models including indicator variables for month-and-year of birth for each individual in the sample instead of the linear year-of-birth trend – the results remain significant and effect sizes are only marginally smaller (Table 4, second column).

Because of the initial and persistent differences with respect to schooling outcomes in one of the *fresco* villages, Espíritu Santo, it is possible that including individuals from this village in the comparison group has led to biases in the estimated effects. To ensure that this is not the case, we re-estimate the models using individuals from small villages only and, separately, from large villages only (dropping in each case the village indicator variable for the now excluded village). These results are shown in Table 4, columns three and four. While less precisely estimated (in part due to smaller sample sizes), all but the effect on Raven's test z-scores in the small village comparison remain significant and all the point estimates are similar, if not larger, than for the combined sample. Indeed, estimates for two of the three outcomes are larger for the large villages (which does not include Espíritu Santo). We conclude that initial differences in Espíritu Santo do not appear to be driving the results.

We next explore altering the comparison groups along a different dimension, whether they were too young or too old to have been fully exposed during their first 36 months of life. Keeping the exposed group the same (those completely exposed from birth to 36 months of age), we use as comparisons only those too young to have been exposed (i.e., born after 1975) and, separately, only those too old to have been exposed (i.e., born before 1969). Again, estimated coefficients (Table 4, columns five and six) are similar in magnitudes to the basic results, though they are not statistically significant for these estimations based on smaller sample sizes.

As discussed in Section 2, the specification of complete exposure from birth to 36 months has the characteristic that it may underestimate the effect of the nutritional intervention because even a child exposed up to 35 months of age is treated as not exposed (i.e., $N_i = N_i^A = 0$). These are our preferred estimates, in part because they are conservative. Moreover, when analysing the distribution of exposure (for example, measured as the fraction of the period from birth to 36 months of age the individual was exposed) it is clear that there are two mass points, at zero and at one, with only about a third of the sample taking on a value in between. Nevertheless, when we use this fractional representation instead of N_i and N_i^A (Table 4, column seven), the findings remain statistically significant and all three point estimates of the effect of *atole*, as expected, are slightly larger than in the base specification.

A final check of the identification strategy we consider is to implement a series of placebo tests, specifications that are parallel to the basic specification but estimate a false treatment. One of these is shown in the final column of Table 4. Using only individuals who were either too young or too old for complete exposure from birth to 36 months (and dropping all the fully exposed individuals), we estimate a model in which we temporarily treat those too old as the 'exposed' group, and those too young as not exposed. Point estimates are far from significant and relatively small compared with

Table 4

OLS Estimates of Impact of Exposure to Atole (Relative to Fresco) from Birth to 36 Months on Educational Outcomes Using HCS (2002–4): Alternative Comparison Groups and Specifications

Outcome	Base (1)	Month & year of birth dummies (2)	Small villages only (3)	Large villages only (4)	Exposed cohort versus those too young (5)	Exposed cohort versus those too old (6)	Fractional exposure period (7)	*Placebo test*: Too old as control for too young (8)
Highest grade completed, women	1.168**	1.102*	1.214*	1.533*	1.120	0.705	1.245*	−0.138
	(0.548)	(0.578)	(0.748)	(0.801)	(0.751)	(0.706)	(0.715)	(0.838)
N	786	786	353	433	485	593	786	494
Highest grade completed, men	−0.432	−0.303	−1.048	0.237	−0.996	−0.458	−0.818	0.629
	(0.534)	(0.689)	(0.836)	(0.790)	(0.824)	(0.713)	(0.752)	(1.211)
N	685	685	291	394	447	525	685	398
Reading comprehension (SIA) z-score	0.277**	0.265**	0.328**	0.293*	0.206	0.160	0.284*	−0.090
	(0.110)	(0.130)	(0.150)	(0.180)	(0.172)	(0.149)	(0.151)	(0.227)
N	1,448	1,448	636	812	917	1,102	1,448	877
Nonverbal cognitive ability (Raven's) z-score	0.238**	0.235*	0.193	0.358**	0.268	0.224	0.327*	0.003
	(0.118)	(0.136)	(0.189)	(0.174)	(0.182)	(0.156)	(0.172)	(0.234)
N	1,448	1,448	636	812	917	1,102	1,448	877

Notes. See notes to Tables 2 and 3 and text for description of analyses.

the base results. In addition, in results not shown we estimated the 'effect' of the intervention on characteristics of the individual, or his or her parents, that could not have been affected by the intervention. These include sex of the child, mother's and father's completed grades of schooling, mother's and father's age when the child was born, whether the parents were married, and the wealth index. For none of these nine outcomes is the coefficient on N_i^A statistically significant. This further justifies our inclusion of these variables as controls to increase the precision of the estimated effects of *atole*.

4.5. *Alternative Exposure Periods*

In principle, the data and design of the original study not only allow exploration of the effect of exposure to *atole* from birth to 36 months but also enable an exploration of whether shorter or longer periods of exposure can have greater effects. While the nutritional literature points to early life as being critical for linear growth, the actual window of growth failure has been shown to vary across populations. Moreover, it is possible that the critical window is different for educational outcomes than for physical growth. We explored the impact of varying N_i and N_i^A across all possible intervals of six months and multiples of six months up to 42 months (not shown), including a consideration of exposure in utero since pregnant mothers also were encouraged to attend the feeding centres. In general, we find that the OLS results reported in Tables 2 and 3 are little changed (both in terms of magnitude of the coefficients and their significance levels) when the exposure periods are altered slightly, for example, considering birth to 30 months, 6 months to 36 months, or 6 months to 42 months. This is not surprising, however, given the high linear correlation among these different exposure variables – due to the fact that many of the same children fall into adjacent categories, indicator variables for adjacent categories differing by only 6 months tend to be correlated at 0.9 and above, those differing by 12 months at 0.8 and above. As a result, it is intractable to separate empirically the effects from different exposure periods for differences of these magnitudes. When we shorten the exposure period to 12 month intervals between zero and 36 months, in general the results are weaker with smaller and less often significant estimates of the effect of *atole*, suggesting that exposure for more than 12 months is important. When we substantially alter the exposure period, for example considering those exposed from age 36 months to 72 months, results for the main outcomes have small estimated coefficients and are insignificant. From this evidence, then, we conclude that it is exposure during the earlier two to three years that are relatively more important than later years, though we are unable to provide strong evidence that the first two years are more important than the third, for example.

4.6. *Sensitivity to Attrition*

The estimates presented above are based on a sample of 1,471 (or 1,448) individuals, 62% (61%) of the original 2,392 sample members. Despite the considerable effort and success in tracing and re-interviewing participants from the original sample, however, attrition of nearly 40% is substantial.[16] Moreover, as shown in Grajeda *et al.* (2005), attrition in the sample is associated with a number of initial conditions, in ways that differ by the reason for attrition (e.g., migration versus failure to interview someone who was located). What is of ultimate concern in this analysis is not the level of attrition, however, but whether, and to what extent, the attrition invalidates the inferences we make using these data.

We address concerns about sample attrition bias in three ways. First, in the specifications already shown, we include a large number of covariates, many of which, in addition to playing a role in educational outcomes, are themselves associated with attrition, including being male (+), expected age in 2002 (−), the wealth index at the time of the intervention (+), and whether the wealth index could not be computed because of missing information (+) (Grajeda *et al.* 2005). Conditional on the maintained assumptions about the correct functional form, attrition selection on right-side variables does not lead to attrition bias (Fitzgerald *et al.*, 1998*b*).

Second, we take advantage of the existence of two other sources of data on these same individuals. The first are village censuses carried out in 1987, 1996 and 2002 (described in Maluccio *et al.* 2005*c*). These village censuses collected completed grades of schooling for all villagers, and thus for all HCS original sample members (still) residing in the village at the time of each census. This means that for individuals exposed to the intervention and measured in the 1987, 1996, or 2002 village census, we have an observation on their completed schooling at that time, regardless of whether or not they were interviewed in HCS. While this measure is potentially censored, for those above 18 (everyone in the 1996 village census) it is likely to be a reliable measure of completed grades for all but a few. The second source of information is the Follow-up 1988–9 survey, which also contains completed grades (measured via individual level surveys) and, in addition, SIA and Raven's test scores, for original sample members living either in the original study villages or in Guatemala City – since unlike HCS, migrants to other parts of the country were not included in that survey (Martorell *et al.*, 1995*a*). As with the HCS sample measured in 2002–4, those interviewed in 1988–9 represent a select sample but, importantly, a different one. In particular, nearly 100 sample members interviewed in the 1988–9 follow-up had emigrated internationally by

[16] A related problem is that of mortality selection (Pitt, 1997; Pitt and Rosenzweig, 1989). Indirect evidence that mortality selection exists in the sample is that higher risk of death is associated with younger ages (those born later) in the complete sample of 2,392. The older sample members represent the survivors of their respective birth cohorts and, hence, they experienced a lower mortality rate (because most mortality was in infancy) compared with the later birth cohorts in the study who were followed from birth. Because the fieldwork began in 1969 and included all children under seven years of age, it excluded all children from the villages born between 1962 and 1969 who died before the start of the survey. Another facet of mortality selection, however, has to do with the intervention itself, which may have decreased mortality rates among the younger cohort in *atole* versus *fresco* villages (Rose *et al.*, 1992). To the extent the variables included in our models are associated with these forms of selection, our estimates partly control for mortality selection, though we do not implement any special methodology to do so. Our identification strategy using a cohort control guards against the mortality selection due to the timing and sample strategy incorporating all children under seven alive in 1969 or born later. It is ambiguous how an effect of *atole* on mortality influences the estimates based on the surviving sample. Arguably it may have kept alive less robust children who would have had lower educational outcomes in life leading to our estimates of the effect of *atole* being downward biased. When we make the alternative extreme assumption of assigning zero grades of education and zero on all test scores to all those children who died before age 7, however, we find effects that are similar to, but slightly smaller in magnitude than, those reported in the article.

2002. To the extent that findings are similar across these different samples, then, it is evidence that attrition bias is not driving the results.

The correspondence between the completed grades measure from the village census surveys and from HCS, for those measured in both data sets, is very high, with a correlation of 0.94 and only 8% of the observations differing by more than one grade of completed schooling. Combining information from the 1987, 1996 and 2002 village censuses, however, yields 2,154 observations, 90% of the original sample (and therefore including many who had migrated by the time of HCS). Using this village census-based sample, the estimated effects of *atole* on grades attained for women are remarkably similar to those based on HCS, including the exact same set of controls (Table 5).

Table 5

Estimates of Impact of Exposure to Atole (Relative to Fresco) from Birth to 36 Months on Educational Outcomes and the Probability of Attrition Using HCS (2002–4), Follow-up (1988–9), and Village Census (2002) Samples

	HCS 2002–4	Combined Village Censuses (1987, 1996 & 2002)	Follow-up 1988–9
Educational Outcomes (OLS)			
Highest grade completed, women	1.168**	1.334***	0.454
	(0.548)	(0.466)	(0.298)
N	786	1,063	652
Highest grade completed, men	−0.432	−0.421	−0.417
	(0.534)	(0.439)	(0.285)
N	685	1,091	671
Reading comprehension (SIA) z-score	0.277**	-	0.200**
	(0.110)		(0.101)
N	1,448		1,368
Nonverbal cognitive ability (Raven's) z-score	0.238**	-	0.188*
	(0.118)		(0.112)
N	1,448		1,368
Age range of respondents at time of data collection	25–42	18–40	11–27
Attrition (probit)			
(1) if in sample			
Women only [$N = 1{,}162$]	0.004	0.021	0.081
	(0.063)	(0.026)	(0.063)
Men only [$N = 1{,}230$]	−0.035	0.024	−0.057
	(0.058)	(0.023)	(0.055)
Women & men combined [$N = 2{,}392$]	−0.015	0.023	0.003
	(0.042)	(0.018)	(0.041)

Notes. For *Educational Outcomes* (top panel), results in column 1 are based on the HCS sample (repeated from Tables 2 and 3). Results in column 2 are based on individuals for whom information was available in a village census survey carried out after the time they were 18 years of age or older. Results in column 3 are based on a sample that consists of all individuals who were exposed to the INCAP supplementation intervention between 1969 and 1977 and who were subsequently re-interviewed in 1988–9. Additional covariates controlled for include all those indicated in Table 2. For *Attrition* (bottom panel), sample consists of all individuals who were exposed to the INCAP supplementation intervention between 1969 and 1977 and the indicator variable is 1 if there were interviewed in the sample considered in each column. Interaction effect on *atole* calculated using methodology proposed by Norton *et al.* (2004). Standard errors for both panels are calculated allowing for clustering at the mother level and reported in parentheses (Statacorp, 2007). * Significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

In Table 5, we also present results based on the data collected in 1988–9, reporting the coefficients on N_i^A for grades completed (for women and men separately) and SIA and Raven's test z-scores (for combined men and women). With the exception of grades completed, the results are again similar to the basic findings. Consistent with findings of Pollitt *et al.* (1993), in 1988–9, there is no evidence of a programme effect on highest grade attained for women, perhaps because this was measured at a relatively young age (about a third age 15 or less) for many of the respondents when they might not have completed their schooling. Coefficient point estimates are smaller for both SIA and Raven's test z-scores, about 70%–80% the size of the estimated effect based on HCS, suggesting the possibility that the effects may be increasing over time.

The final assessment of attrition we present is from estimation of attrition probits conditioning on the same set of variables and determining whether the intervention had a large or statistically significant effect on the probability of attrition. In the bottom panel of Table 5, we present the estimated marginal effects from a series of probits predicting whether the observation is in the sample for each of these three samples and separately for the women only, men only, and combined samples – in all cases, the estimated effects are small and insignificant.[17] The estimates based on alternative samples with different selection processes, as well as the attrition probits, point to the same conclusion: the results described above are not driven by attrition bias.

4.7. *The Role of Energy and Protein*

In this Section, we explore the magnitudes of the effect of improved nutrition by examining how different caloric and protein intakes from the supplements affected those same outcomes. We endogenise the directly measured individual-level data on caloric and protein intakes from supplements in relations estimating the same set of educational outcomes. Specifically, we replace N_i^A in relation (1) with average daily kilocalories (and, in separate models, average daily protein in grams) consumed from either *atole* or *fresco* supplement from birth to 36 months of age.[18] We treat this as an endogenous variable and use N_i^A as an instrument estimating via two-stage least squares (2SLS), yielding estimates of the treatment-on-the-treated.

In the first (third) column of Table 6, we present the first-stage results for the excluded instrument in the top panel and the 2SLS estimated coefficients on calories (protein) from supplement in the second panel. For both calories and protein, the

[17] In addition, we consider two other assessments of attrition. First, we re-weight estimates based on the methodology proposed by Fitzgerald *et al.* (1998*a*, *b*) and find nearly identical results to those presented above. Second, we compare nutritional outcomes measured in the 1970s for those eventually re-interviewed in HCS and those not re-interviewed. Average height-for-age measured at 48 months of age and 72 months of age are remarkably similar between those who attrited and those who did not. Height-for-age z-scores for the two groups are all within 0.006 of one another (mean for 48-month olds is −2.074, SD 1.03 and mean for 72-month olds is −1.666, SD 1.028) and the lowest p-value on a t-test with unequal variances is p = 0.91. There does not appear to be any obvious selection between those interviewed or not based on early-life nutritional status.

[18] Because *fresco* had calories but no protein and *atole* had both but, of course, in fixed proportion, calories and protein consumed from supplement from birth to 36 months were correlated above 0.98, so it is not statistically feasible to include both in the same relation.

Table 6

2SLS Estimates of Impact of Average Daily Calories or Protein from Supplement from Birth to 36 Months on Educational Outcomes Using HCS (2002–4)

	Calories/day from supplement		Protein (grams/day) from supplement	
First-stage estimates				
Coefficient on exposure to *atole* relative to *fresco* from birth to 36 months	9.835***		0.290***	
	(1.946)		(0.110)	
F-test (on excluded instrument)	24.4		31.8	
Second stage estimates	2SLS estimate	Implied impact of exposure to *atole*	2SLS estimate	Implied impact of exposure to *atole*
Highest grade completed, women				
	0.032*	2.51	0.427*	2.391
	(0.017)	[0.32, 4.71]	(0.224)	[0.34, 4.45]
Highest grade completed, men				
	−0.017	−1.39	−0.201	−1.123
	(0.022)	[−4.23, 1.44]	(0.245)	[−3.38, 1.14]
Reading comprehension (SIA) z-score				
	0.009**	0.713	0.113**	0.631
	(0.004)	[0.21, 1.22]	(0.048)	[0.19, 1.07]
Nonverbal cognitive ability (Raven's) z-score				
	0.008*	0.612	0.097*	0.542
	(0.004)	[0.09, 1.14]	(0.050)	[0.08, 1.00]

Notes. Endogenous variable in equations exploring the role of calories (protein) is average daily calories (average daily grams of protein) from supplement from birth to 36 months, instrumented with N_i^A. Coefficient on the excluded instrument (N_i^A) and related statistics for first-stage regression predicting average daily calories (average daily grams of protein) shown in first panel. Two-stage least squares regressions include all other covariates listed in Table 2 (including N_i). Standard errors are calculated allowing for clustering at the mother level and reported in parentheses (Statacorp, 2007). * Significant at the 10% level; ** significant at the 5% level; *** significant at 1% level. The implied impact is the estimated coefficient on an additional calorie (gram of protein) × the average daily gain in calories, 79.8 kcal, (protein, 5.6 grams) for those exposed to *atole* from birth to 36 months. 90% confidence intervals of the implied impact shown in brackets.

first-stage regressions have high predictive power, with the F-statistics on the excluded instrument for this exactly identified relationship of 24 or higher (Bound *et al.*, 1995).[19] Turning to the 2SLS results in the bottom panel, the three principal findings are all statistically significant at a 10% level or better (and the effect of the intervention on highest grade completed for men remains insignificant). To gauge the magnitude of

[19] We also consider distance to the feeding centre as another possible instrument. However, some of the feeding centres were located close to primary schools and, as such, they might be expected to have a direct effect on outcomes, particularly grade attainment and SIA scores. This hypothesis is supported by the rejection of the Hansen J overidentification test (whose null hypothesis is that the overidentifying restrictions are valid, i.e., that the instruments do not belong in the second-stage equation) (Hayashi, 2000) for some outcomes, particularly SIA tests, when we include distance to feeding centre indicator variables as additional overidentifying instruments.

these treatment effects for calories, in column 2 we multiply the estimated parameter reported in column 1 by 79.8 kcal, the mean daily intake of calories from supplement consumed by individuals exposed to *atole* from birth to 36 months, yielding the average impact of exposure to *atole* on these outcomes. We also report, in square brackets, the associated 90% confidence intervals. In the fourth column we carry out the same calculation for grams of protein per day from *atole* (which averaged 5.6 grams).

Because the 2SLS estimates capture the effect of the treatment-on-the-treated, we expect these estimated impacts to be larger than the intent-to-treat reduced form results presented above. Table 6 shows that in all cases, the implied impacts are approximately twice the size of the reduced form results, suggesting the intent-to-treat estimates are relatively conservative. Nevertheless, the reduced form estimates lie comfortably within the 90% confidence intervals of the 2SLS estimates.[20] The implied impacts are of similar magnitude for both calories and protein. In conjunction with the finding that exposure to *atole* from birth to 36 months led to greater total consumption of calories and protein (Islam and Hoddinott, 2009), these results support the interpretation of our findings that the intervention led to substantially improved educational outcomes via improved nutrient intakes. Since we cannot, in this study, distinguish between the effect of energy or other nutrients, our interpretation is that the results demonstrate the value of increasing nutrient intakes and dietary quality in general.

5. Conclusion

We estimate the impact of a community-level nutritional intervention in which highly nutritious supplements were provided, on several different measures related to education over the life cycle. We advance beyond the previous literature by using longitudinal data from a nutritional experiment carried out in rural Guatemala from 1969–77 that enable us to consider educational measures not only for school-age individuals but also for prime-age adults. The results of our intent-to-treat estimates show significantly positive, and fairly substantial, effects of exposure to the nutritional intervention from birth to 36 months of age (but not later) a full 25 years after the intervention ended. These include higher completed grades by women (1.2 grades) (but no effect for men); higher scores on reading comprehension tests for both women and men (one quarter of a SD); and higher scores on non-verbal cognitive ability tests for both women and men (one quarter of a SD). The findings are robust to variations in the identification strategy, slight changes in how we define the exposure period and to sample attrition. Two-stage least squares results that endogenise the actual supplement intakes measured during the intervention corroborate the magnitudes of these effects.

The results provide the first evidence of its kind from a prospective survey of the important role played by pre-school nutrition in subsequent adult educational attainments and thus underscore the value of a life cycle approach to education that includes both the pre-school and the adult periods. Interventions that include nutritional supplements to very young children, or in other ways improve their nutrient intakes,

[20] Results are qualitatively similar when we instead consider educational outcomes measured in the Follow-up 1988–9 (not shown).

can have substantial, long-term educational consequences. Related research with these data has demonstrated that improvements in nutrition also led to greater economic productivity, as reflected in higher wages for men (Hoddinott *et al.*, 2008). While a full cost–benefit accounting is beyond the scope of this article, rough calculations indicate that the direct costs of the average intake per child of *atole* for three years in the 1970s was on the order of $15. In present day Guatemala, the same amount of a similar *atole*, commercially marketed *Incaparina*, would cost about four times that. The substantial long-term effects reported here, in conjunction with relatively modest costs, provide strong justification for such early childhood interventions.

Middlebury College
International Food Policy Research Institute
University of Pennsylvania
Emory University
International Food Policy Research Institute
Emory University

Submitted:19 January 2007
Accepted: 24 June 2008

References

Alderman, H., Behrman, J.R., Kohler, H.P., Maluccio, J.A. and Cotts Watkins, S. (2001*a*). 'Attrition in longitudinal household survey data: some tests for three developing country samples', *Demographic Research [Online]*, vol. 5(4), pp. 79–124, available at <http://www.demographic-research.org>.

Alderman, H., Behrman, J.R., Lavy, V. and Menon, R. (2001*b*). 'Child health and school enrollment: a longitudinal analysis', *Journal of Human Resources*, vol. 36(1), pp. 185–205.

Alderman, H., Behrman, J.R., Ross, D. and Sabot, R.H. (1996). 'The returns to endogenous human capital in Pakistan's rural wage labor market', *Oxford Bulletin of Economics and Statistics*, vol. 58(1), pp. 29–55.

Alderman, H., Hoddinott, J. and Kinsey, B. (2006). 'Long-term consequences of early childhood malnutrition', *Oxford Economic Papers*, vol. 58(3), pp. 450–74.

Angrist, J.D. and Lavy, V. (2002). 'The effect of high school matriculation awards: evidence from randomised trails', NBER Working Paper No. 9389.

Baker, A.J.L., Piotrkowski, C.S. and Brooks-Gunn, J. (1998). 'The effects of home instruction program for preschool youngsters (HIPPY) on children's school performance at the end of the program and one year later', *Early Childhood Research Quarterly*, vol. 13(4), pp. 571–88.

Behrman, J.R. (1996). 'Impact of health and nutrition on education', *World Bank Research Observer*, vol. 11(1), pp. 23–37.

Behrman, J.R., Cheng, Y. and Todd, P. (2004). 'Evaluating preschool programs when length of exposure to the program varies: a nonparametric approach', *Review of Economics and Statistics*, vol. 86(1), pp. 108–32.

Behrman, J.R., Hoddinott, J., Maluccio, J.A., Soler-Hampejsek, E., Behrman, E.L., Martorell, R., Quisumbing, R.L., Ramírez, M. and Stein, A.D. (2008). 'What determines adult skills? Impacts of preschool, school-years and post-school experiences in Guatemala', University of Pennsylvania, Working Paper.

Bergeron, G. (1992). 'Social and economic development in four ladino communities of Eastern Guatemala: a comparative description', *Food and Nutrition Bulletin*, vol. 14(3), pp. 221–36.

Bertrand, M., Duflo, E. and Mullainathan, S. (2004). 'How much should we trust differences-in-differences estimates?', *Quarterly Journal of Economics*, vol. 119(1), pp. 249–75.

Boissiere, M., Knight, J.B. and Sabot, R.H. (1985). 'Earnings, schooling, ability and cognitive skills', *American Economic Review*, vol. 75(5), pp. 1016–30.

Bound, J., Jaeger, D.A. and Baker, R.M. (1995). 'Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak', *Journal of the American Statistical Association*, vol. 90(430), pp. 443–50.

Cunha, F., Heckman, J.J., Lochner, L. and Masterov, D.V. (2008). 'Interpreting the evidence on life cycle skill formation', in (E. Hanushek and F. Welch, eds.), *Handbook of the Economics of Education*, pp. 697–812, Amsterdam: North Holland.

Currie, J. and Thomas, D. (1995). 'Does Head Start make a difference?', *American Economic Review*, vol. 85(3), pp. 341–64.

Donald, S.G. and Lang, K. (2007). 'Inference with difference-in-differences and other panel data', *Review of Economics and Statistics*, vol. 89(2), pp. 221–33.

Estudio 1360. (2002). 'Changes in the socioeconomic and cultural conditions that affect the formation of human capital and economic productivity', final report presented to The Institute of Nutrition of Central America and Panama (INCAP), May 13, Photocopy.

Fitzgerald, J.M., Gottschalk, P. and Moffitt, R.A. (1998*a*). 'An analysis of sample attrition in panel data', *Journal of Human Resources*, vol. 33(2), pp. 251–99.

Fitzgerald, J.M., Gottschalk, P. and Moffitt, R.A. (1998*b*). 'The impact of attrition in the PSID on intergenerational analysis', *Journal of Human Resources*, vol. 33(2), pp. 300–44.

Garces, E., Thomas, D. and Currie, J. (2002). 'Longer-term effects of Head Start', *American Economic Review*, vol. 92(4), pp. 999–1012.

Glewwe, P., Jacoby, H. and King, E.M. (2000). 'Early childhood nutrition and academic achievement: a longitudinal analysis', *Journal of Public Economics*, vol. 81(3), pp. 345–68.

Glewwe, P. and King, E.M. (2001). 'The impact of early childhood nutrition status on cognitive achievement: does the timing of malnutrition matter?', *World Bank Economic Review*, vol. 15(1), pp. 81–114.

Grajeda, R., Behrman, J.R., Flores, R., Maluccio, J.A., Martorell, R. and Stein, A.D. (2005). 'The human capital study 2000–04: tracking, data collection, coverage, and attrition', *Food and Nutrition Bulletin*, vol. 26(2) (Supplement 1), pp. S15–23.

Grantham-McGregor, S.M., Cheung, Y.B., Cueto, S., Glewwe, P., Richter, L. and Strupp, B. (2007). 'Developmental potential in the first 5 years for children in developing countries', *Lancet*, vol. 369(9555, 6 January), pp. 60–70.

Grantham-McGregor, S.M., Fernald, L.C. and Sethuraman, K. (1999*a*.) 'Effects of health and nutrition on cognitive and behavioral development in children in the first three years of life: Part 1: low birth weight, breastfeeding and protein-energy malnutrition', *Food and Nutrition Bulletin*, vol. 20(1), pp. 53–75.

Grantham-McGregor, S.M., Fernald, L.C. and Sethuraman, K. (1999*b*). 'Effects of health and nutrition on cognitive and behavioral development in children in the first three years of life: Part 2: infections and micronutrient deficiencies: iodine, iron, and zinc', *Food and Nutrition Bulletin*, vol. 20(1), pp. 76–99.

Grantham-McGregor, S.M., Walker, S., Chang, S. and Powell, C. (1997). 'Effects of early childhood supplementation with and without stimulation on later development in stunted Jamaican children', *American Journal of Clinical Nutrition*, vol. 66(2), pp. 247–53.

Habicht, J.P. and Martorell, R. (1992). 'Objectives, research design, and implementation of the INCAP longitudinal study', *Food and Nutrition Bulletin*, vol. 14(3), pp. 176–90.

Habicht, J.P., Martorell, R. and Rivera, J.A. (1995). 'Nutritional impact of supplementation in the INCAP longitudinal study: analytic strategies and inferences', *Journal of Nutrition*, vol. 125(4S), pp. 1042S–50S.

Habicht J.P., Schwedes, J.A., Arroyave, A. and Klein, R.E. (1973). 'Biochemical indices of nutrition reflecting ingestion of a high protein supplement in rural Guatemalan children', *American Journal of Clinical Nutrition*, vol. 26(10), pp. 1046–52.

Haddad, L., Alderman, H., Appleton, S., Song, L. and Yohannes, Y. (2003). 'Reducing child malnutrition: how far does income growth take us?', *World Bank Economic Review*, vol. 17(1), pp. 107–31.

Harcourt Assessment. (2008). Raven's progressive matrices, available at <http://www.harcourt-uk.com/product.aspx?n=1342&s=1346&skey=3068>, accessed February 26, 2008.

Hayashi, F. (2000). *Econometrics*, Princeton, NJ: Princeton University Press.

Hoddinott, J., Maluccio, J.A., Behrman, J.A., Flores, R. and Martorell, R. (2008). 'Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults', *Lancet*, vol. 371(9610, 2 February), pp. 411–6.

Islam, M. and Hoddinott, J. (2009). 'Evidence of intra-household flypaper effects from a nutrition intervention in rural Guatemala', *Economic Development and Cultural Change*.

Johnston, F., Low, S., de Baessa, Y. and MacVean, R. (1987). 'Interaction of nutritional and socioeconomic status as determinants of cognitive achievement in disadvantaged urban Guatemalan children', *American Journal of Physical Anthropology*, vol. 73(4), pp. 501–6.

Li, H. Barnhart, H.X., Stein, A.D. and Martorell, R. (2003). 'Effects of early childhood supplementation on the educational achievement of women', *Pediatrics*, vol. 112(5), pp. 1156–62.

Li, H., DiGirolamo, A., Barnhart, H.X., Stein, A.D. and Martorell, R. (2004). 'Relative importance of birth size and postnatal growth for women's educational achievement', *Early Human Development*, vol. 76(1), pp. 1–16.

Maluccio, J.A., Martorell, R. and Ramírez, L.F. (2005*a*). 'Household expenditures and wealth among young Guatemalan adults', *Food and Nutrition Bulletin*, vol. 26(2) (Supplement 1), pp. S110–9.

Maluccio, J.A., Murphy, A. and Yount, K.M. (2005*b*). 'Research note: a socioeconomic index for the INCAP Longitudinal Study 1969–77', *Food and Nutrition Bulletin*, vol. 26(2) (Supplement 1), pp. S120–4.

Maluccio, J.A., Melgar, P., Mndez, H., Murphy, A. and Yount, K.M. (2005*c*). 'Social and economic development and change in four Guatemalan villages: demographics, schooling, occupation, and assets', *Food and Nutrition Bulletin*, vol. 26(2) (Supplement 1), pp. S55–67.

Manuel, H. T. (1967). *Technical Reports, Tests of General Ability and Tests of Reading, Interamerican Series*, San Antonio, TX: Guidance Testing Associates.

Martorell, R. (1992). 'Overview of long-term nutrition intervention studies in Guatemala', *Food and Nutrition Bulletin*, vol. 14(3), pp. 270–7.

Martorell, R. (1997). 'Undernutrition during pregnancy and early childhood and its consequences for cognitive and behavioral development', in (M.E. Young ed.), *Early Child Development: Investing in our Children's Future*, pp. 39–83, Amsterdam: Elsevier.

Martorell, R., Behrman, J.R., Flores, R. and Stein, A.D. (2005). 'Rationale for a follow-up study focusing on economic productivity', *Food and Nutrition Bulletin*, vol. 26(2) (Supplement 1), pp. S5–14.

Martorell, R., Habicht, J.P. and Rivera, J.A. (1995*a*). 'History and design of the INCAP longitudinal study (1969–77) and its follow up (1988–89)', *Journal of Nutrition*, vol. 125(4S), pp. 1027–41S.

Martorell, R., Schroeder, D.G., Rivera, J.A. and Kaplowitz, H.J. (1995*b*). 'Patterns of linear growth in rural Guatemalan adolescents and children', *Journal of Nutrition*, vol. 125(4S), pp. 1060–7S.

Moulton, B.R. (1990). 'An illustration of a pitfall in estimating the effects of aggregate variables on micro units', *Review of Economics and Statistics*, vol. 72(2), pp. 334–8.

Norton, E.C., Wang, H. and Ai, C. (2004). 'Computing interaction effects and standard errors in logit and probit models', *Stata Journal*, vol. 4(2), pp. 103–16.

Pitt, M.M. (1997). 'Estimating the determinants of child health when fertility and mortality are selective', *Journal of Human Resources*, vol. 32(1), pp. 127–58.

Pitt, M.M. and Rosenzweig, M.R. (1989). 'The selectivity of fertility and the determinants of human capital investments: parametric and semi-parametric estimates', Brown University Department of Economics Working Paper 89–30.

Pivaral, V.M. (1972). *Características Económicas y Socioculturales de Cuatro Aldeas Ladinas de Guatemala*, Ministerio de Educación Pública, Instituto Indigenista Nacional, Guatemala: INCAP.

Pollitt, E., Gorman, K.S., Engle, P., Martorell, R. and Rivera, J.A. (1993). 'Early supplementary feeding and cognition: effects over two decades', *Monographs of the Society for Research in Child Development*, Serial No. 235, 58 (7).

Ramey, C.T., Campbell, F.A. and Blair, C. (1998). 'Enhancing the life course for high-risk children', in (J. Crane, ed.), *Social Programs that Work*, pp. 184–99, New York: Russell Sage Foundation.

Raven, J.C., Court, J.H. and Raven, J. (1984). *Manual for Raven's Progressive Matrices and Vocabulary Scales, Section 2: Coloured Progressive Matrices*, London: H. K. Lewis.

Read, M.S. and Habicht, J.P. (1992). 'History of the INCAP longitudinal study on the effects of early nutrition supplementation in child growth and development', *Food and Nutrition Bulletin*, vol. 14(3), pp. 169–75.

Rivera, J.A., Martorell, R., Ruel, M.T., Habicht, J.P. and Haas, J.D. (1995). 'Nutritional supplementation during preschool years influences body size and composition of Guatemalan adolescents', *Journal of Nutrition*, vol. 125(4S), pp. 1068–77S.

Rose, D., Martorell, R. and Rivera, J.A. (1992). 'Infant mortality rates before, during, and after a nutrition and health intervention in rural Guatemalan villages', *Food and Nutrition Bulletin*, vol. 14(3), pp. 215–20.

Schroeder, D.G., Kaplowitz, H.J. and Martorell, R. (1992). 'Patterns and predictors of participation and consumption of supplement in an intervention study in rural Guatemala', *Food and Nutrition Bulletin*, vol. 14(3), pp. 191–200.

Schroeder, D.G., Martorell, R., Rivera, J.A., Ruel, M.T. and Habicht, J.P. (1995). 'Age differences in the impact of nutritional supplementation on growth', *Journal of Nutrition*, vol. 125(4S), pp. 1051–9S.

Schweinhart, L.J. and Weikart, D.P. (1998). 'High/Scope Perry Preschool Program effects at age twenty-seven', in (J. Crane ed.), *Social Programs that Work*, pp. 148–62, New York: Russell Sage Foundation.

StataCorp. (2007). *Stata Statistical Software: Release 10.0*, College Station, TX: Stata Corporation.

Stein, A.D., Behrman, J.R., DiGirolamo, A., Grajeda, R., Martorell, R., Quisumbing, A.R. and Ramakrishnan, U. (2005). 'Schooling, educational achievement and cognitive functioning among young Guatemalan adults', *Food and Nutrition Bulletin*, vol. 26(2) (Supplement 1), pp. S46–54.

Stein, A.D., Wang, M., DiGirolamo, A., Grajeda, R., Ramakrishnan, U., Ramírez-Zea, M., Yount, K.M., and Martorell, R. (2008). 'Nutritional supplementation in early childhood, schooling and intellectual functioning in adulthood: a prospective study in Guatemala', *Archives of Pediatric Adolescent Medicine*, vol. 162(7) (July), pp. 612–18.

UNICEF (United Nations Children's Fund). (1998). *The State of the World's Children 1998*, New York, Oxford University Press.

Victora, C.V., Adair, L., Fall, C., Hallal, P.C., Martorell, R., Richter, L. and Sachdev, H.S. (2008). 'Maternal and child undernutrition: consequences for adult health and human capital', *Lancet*, vol. 371(9609, 26 January), pp. 340–57.

Wooldridge, J.M. (2003). 'Cluster-sample methods in applied econometrics', *American Economic Review*, vol. 93(2), pp. 133–8.

World Bank. (2001). *Engendering Development*, World Bank Policy Research Report, Washington DC, World Bank and Oxford University Press.

World Bank. (2003). *Poverty in Guatemala*, Report No. 24221-GU, Washington, DC: World Bank.