Session 2D: Inequality and Growth in Developing Countries
Time: Monday, August 6, 2012 PM

Paper Prepared for the 32nd General Conference of
The International Association for Research in Income and Wealth

Boston, USA, August 5-11, 2012

Long-Term Impact of Malnutrition on Education Outcomes for Children in Rural Tanzania

Lucia Luzi

For additional information please contact:

Name: Lucia Luzi
Affiliation: University of Venice

Email Address: lucia.luzi@unive.it

Long-Term Impact of Malnutrition on Education Outcomes for Children in Rural Tanzania

Lucia Luzi[a]

[a] *Ca' Foscari University of Venice, San Giobbe, Cannaregio 873, 30121, Venice, Italy*

E-mail address: lucia.luzi@unive.it

May 2012

JEL Classification Code: I15, I25

Abstract

This paper investigates the long-term impact of early childhood malnutrition, in children living in a rural area of Tanzania, on their subsequent educational achievements as young adults. The data used are of an exclusive long term panel data set collected in the Kagera Health and Development Survey. Infants born in the early Nineties were traced and interviewed in 2004. To perform the main objective of the work, any attrition due to household or environmental characteristics is removed by differencing among siblings. Additionally, a broad investigation on the weather conditions that prevailed during infancy is conducted, in order to attain the instruments to face the existing endogeneity proper of the health variable.

Estimation results show that malnutrition and poor health experienced during early life have long term effects on the child's human capital growth. Specifically, improving the Tanzanian child's health status would result in an additional 28% probability of completing primary education. This result emerges when the two districts of Kagera - where the refugees escaped from the genocides of Burundi and Rwanda in the early Nineties - are excluded from the analysis.

Key words: Primary Education, Child Health and Nutrition, Weather Shocks, Fixed Effects, Instrumental Variables, Tanzania.

1. Introduction

Persons and households, principally those located in rural areas of developing countries, are often exposed and are vulnerable to exogenous shocks, like weather-related shocks. These kinds of episodes drastically affect household income and thus, the well-being of its components[1]. Such events can also produce variations in consumption, with consequent decreases of utility of both persons and households.

In the context of a developing country, a body of literature investigates the relationship between infants' and preschoolers' nutritional status and adolescents' educational performance. [2] It is reasonable to assume that enhanced health leads to further education and higher achievements. Students in good health miss fewer days of school and, thus, learn more. Such healthier students and later, adults are efficient learners of new skills and competencies; consequently, they invest more in schooling. If past health is an input into current health status, this causal path may cause permanent effects. Conversely, malnourished children are not as productive and efficient as their well nourished peers are . Chronic malnutrition experienced during early life inhibits growth, retards mental development, reduces motivation and energy level, causing a reduced educational attainments and delayed entry into schools. (Pollitt, 1984). Furthermore, malnourished adults cannot work as efficiently as their well-nourished peers, with major effects on both personal and national income.

External shocks may cause permanent damage to a child's future welfare and cognitive abilities (World Bank, 1993); hence, further empirical investigation is required to precisely assess the magnitude of the effect of such shocks on early child growth.

From a policy point of view, understanding the relationship between health and education is crucial to formulate new possible policy interventions targeted to improve children's health status in developing countries. Nevertheless, investigations on this topic confirm that a clear articulation of health and nutrition issues is largely absent from the education policy agenda in many developing

[1] A growing literature analyses both households kinds of respond to these shocks and the effectiveness of these responds in reducing fluctuations in consumption (Hoddinott and Kinsey, 2001)

[2] Understanding the nature of the causal relationship between health and education is crucial to determine the exact relation between them. Associations do not necessarily indicate causality. As a matter of fact, estimates are usually likely to be biased in one direction or the other because of unobserved factors, such as parental preferences for health and education. These unobserved factors may determine both nutrition and education outcomes, generating a correlation that is not necessarily causal.

countries. Given the growing importance of the potentially strong connection between nutrition and schooling in developing countries, the analysis presented in this work makes progress in sorting out such a casual relationship.

This paper widens the literature on the determinants of human capital formation; it goes beyond any previous works on the impact of child health and nutrition on their education attainment in developing countries by proposing a specific study.[3] It performs micro-econometric analysis examining the long-term effects of health status presented by individuals, during early childhood, on their subsequent education attainment as young adults. The data used are an exclusive long term panel data set collected for a rural area of the United Republic of Tanzania - the Kagera Health and Development Survey (KHDS) - which covers a period of 13 years, from 1991 to 2004. This paper considers infants born in the early Nineties who were traced and interviewed in 2004. This paper is different from much of the literature on shocks as it assesses the impact at the individual rather than at the household level.

To perform the main objective of the work, any attrition due to household or environmental characteristics is removed by differencing among siblings. Additionally, a broad investigation on weather conditions during infancy is conducted, in order to attain the instruments to face the existing endogeneity proper of the health variable. Rainfall data existing from 1980 to 2004 are used to construct an *ad hoc* indicator, built up by linking historical rainfall for year and location of birth for each person with outcomes of the same individuals as young adults. Specifically, exposure to transitory weather shocks experienced by Kageran children during their birth year is assessed to identify differences in height-for-age of siblings as pre-schoolers, and then, connected to schooling outcomes during early adulthood.

By comparing the anthropometric measures of a Tanzanian preschooler with those of a child in a wealthy reference country, estimation results show that malnutrition and poor health experienced during early life have long term effects on a child's human capital growth. Specifically, improving a child's health status would result in an additional 28% probability of completing primary education.

[3] See, among the others, works of Glewwe and Jacoby (1995), Behrman (1996), Alderman et al. (2001), Glewwe et al. (2001), Maluccio et al. (2006), Alderman, Hoddinott and Kinsey (2006), Fletcher and Lehrer (2009).

This result emerges when the two districts of Kagera, where the refugees escaped from the genocides of Burundi and Rwanda in the early Nineties, are excluded from the analysis.

Section 2 of this paper presents the econometric problem and the identification strategy. Section 3 contains the description of the data set employed and its framework. Section 4 describes the estimation strategy and sample investigation. Section 5 discusses the main findings. Section 6 presents the conclusions and policy implications.

2. Econometric problem and identification strategy

The optimal quantity of investments in knowledge capital and the number of years of formal school completed should be positive functions of the efficiency with which individuals transform all the inputs into their knowledge stock. Efficiency in producing human capital through education is given by factors such as health, physical and mental abilities. Furthermore, investments in health are expected to have positive effects on education since the returns from investment in education last for many periods and health status is positively correlated with life expectancy. Additionally, the opportunity to finance human capital investments should determine the quantity of schooling achieved by the child.

Health and education are partly determined by genetic endowment, but they are not completely exogenous in a life cycle model, since family background and socio-economic environment also play crucial roles. Consequently, investments in a child’s human capital are subject to the constraints imposed by family resources and options available in the community to the individual. It is assumed that in response to signals of their children’s health and ability endowments, altruistic parents opt for inputs that maximize the family’s indirect utility function, subject to a budget constraint made by the production functions for academic achievement and for a child’s health. Some of these inputs enter both schooling and health production functions, creating stocks of human capital for each child in the household. Parents provide their children who have various abilities and health outcomes with different inputs, whether the marginal returns to investments in education of one child is equated, in equilibrium, to the marginal returns to investments in health in their siblings (Fletcher, Lehrer, 2009).

To explain clearly the specific study of this paper, the analysis is based on a two time period model, each period corresponds to the child's age.[4] The first period (t=1) covers the time during which the individual is a newborn or a preschooler. In the second period (t=2) the individual is an adolescent or a young adult. In each period parents make fundamental decisions on their child's human capital investments, with those in the first period being among the most important for a child's developing path. The choice of distinguishing between these two specific periods follows the medical assumptions that the first three years of life are crucial for both cognitive and physical development for an individual. For this reason, what happens during this period of life has long-term effects. The manifestation of health and nutritional shocks occurs long before parental investments in human capital are completed.

Following the well known economic model of human capital investment (Rosenzweig and Wolpin, 1988), the central interest of the analysis is represented by a linear function for achievements in education by the child i in the second time period. This function can be indicated by the vector S_{i2} as follows:

(1) $S_{i2} = \alpha_H f(H_{i1}) + \alpha_{C2}\, g(C_{i2}) + \varepsilon_{i2}$

In equation (1) presented above, $f(H_{i1})$ is a function of health and nutritional status of the respondent interviewed during the first time period and reveals parental decisions on investing in their child's health. Health variable has direct and positive impacts on education outcome through several paths. It may influence the child's physical energy level, which determines the time necessary for learning and the school attendance. Moreover, it affects mental status of the child, having a direct impact on performance at school. Finally, it may affect parents', teachers' or peers' actions towards the child, partly influencing the educational environment.

In the same equation, $g(C_{i2})$ is a vector of observable inputs - individual, household and community characteristics - that affect academic performance. The individual characteristics are the time-invariant demographic variables, such as gender and age. The household background characteristics can be represented by the level of education of both parents, the household size and

[4] This argument follows previous elaborations made by Glewwe et al. (2001), Alderman et al. (2001), Alderman, Hoddinott and Kinsey (2006), Glewwe and Miguel (2008).

composition, the general household environment in which the child grows up etc. Intergenerational transfer of education plays a vital role in households' living standards, maternal education in particular[5]. The community level characteristics include, social culture and values, public infrastructures and programs etc. It also includes the supply of education, which means the level and the quality of schooling, i.e., educational infrastructures, teachers' competences, the quantity of school places available, the pupil-teacher ratio, the adequacy of access. Moreover, it may include the local health environment in which the child lives, air and water quality, as there would be a possibility to contract an infectious disease.

Finally, ε_{i2} represents the individual specific error term that affects the educational outcome of interest and is composed by three elements: the time invariant individual specific characteristics that are unknown by parents during the first years of a child's life (as child's innate ability, motivation and genetic endowment); the time invariant household environment characteristics (as parental preferences on human capital investments of their children, the time spent and care given to cognitive development of their children, the household intellectual atmosphere) common to all children in the household and that influence education; a white noise error term.

To resume, equation (1) describes the role of a child's health in determining academic outcomes. In other words, it illustrates how, *ceteris paribus*, variations in a child's health status could affect performances at school. Each right hand variable (which can be both exogenous and endogenous) has positive and direct impact on S_{i2}, building up a structural relationship.

The child's health status in the first period can be represented as the reduced form in equation (2), which presents a function depending only on exogenous factors:

(2) $$H_{i1} = a_{C1}\, g(C_{i1}) + \varepsilon_{i1}$$

[5] There is a quite extensive socioeconomic literature showing that there are associations between parental family background and early childhood outcomes in developing countries, much of which is reviewed in Behrman and Deolalikar (1988), Strauss and Thomas (1995, 1998). While early childhood outcomes are usually represented by physical health measures, parental family background are generally represented by schooling, occupation, income. A much smaller set of studies explore such relations while attempting to control for familiar aspects of unobserved endowments by using data on siblings. Behrman (1988), for example, uses sibling data on child anthropometrics and nutrient intakes for siblings from rural semi-arid tropical India to estimate to what extent households allocate nutrients among children consistent with parental preferences. The estimates suggest that parental preferences among children are not neutral, but favour to differing degrees across seasons the children who are better off.

H_{i1} is a function of a vector of several observable inputs – indicated by $g(C_{i1})$ – at individual, household and community level that influence investments in health. Moreover, ε_{i1} is the individual specific error term that affects health status and is composed by three elements: the time invariant individual specific characteristics unknown by parents during the first years of a child's life (as child's genetic endowment and inherent physical condition); the time invariant household environment characteristics (as parental preferences for their children's health, the care given to their children, the amount and quality of food given to their child, their ability and sensibility, their involvement in a child's educational growth) common to all children in the household and that influence health status; a white noise error term.

Research on the causal effect of a child's nutritional and health status on schooling, faces many econometric challenges. As Behrman (1996) and Glewwe (2005) note, despite the extensive literature which show positive associations between health problems and school performance, researches on this issue are complex and ambiguous, in part because of a lack of high-quality data with which to address the issue, and in part because there are many possible sources of bias: omitted variable bias from a variable that is correlated with the exogenous variable but is unobserved, so cannot be included in the regression; measurement error bias; sample selection bias; and simultaneous causality bias. In the following sections of the paper, some techniques used to face such estimation problems are discussed.

If complete data on all the right hand variables in equation (1) were available, the function could be estimated by using relatively simple methods (such as *naïve* Ordinary Least Squares), which would produce unbiased estimates of the direct impacts of each variables on academic skills. Unfortunately, in some cases as the present one, it is not possible to have all the required data and some factors, like ability, remain unobserved, causing severe estimation problems.

The stochastic specification shows the problem of spurious correlation between health status and educational achievement. Since a child's health at the first time period is probably correlated with individual or household characteristics included in the error term, the naïve estimate regression of equation (1) would probably be biased. In performing such analysis, an endogeneity problem exists; the health status during infancy can be not orthogonal to the error term, which means that $E(H_{i1}\varepsilon_{i2})\neq 0$. Thus, estimates obtained by using OLS would be either upward or downward biased. This can be driven by several reasons. At first, there is a possible correlation of individual effects. For instance, a child with a poor genetic endowment would probably die between the two periods, which leave a selected sample of people with a better genetic development potential. On the contrary, a child with high genetic development potential would be taller than his contemporaries in both time periods. Another option is the correlation of household effects. As soon as a child's talent or motivation for learning is recognized by parents, it may influence their allocations of health and nutritional inputs. For instance, parents who have a short child in the first period may decide to drive more resources on that child, or otherwise support educational inputs, presuming that she would perform better in intellectual work during the following period. Those genetic characteristics or parental investments are also unobserved to the data analyst.

A possible path to solve the problem related to endogeneity of the childhood nutritional status could be using H_{0i}, the nutritional status of the child in the prenatal life and the first two years of life, which represents the determinant period of child's cognitive development. Employing H_{0i}, instead of H_{i1} is functional, given that the former is uncorrelated with the error term ε_{i2}. Unfortunately, data on health during this period are often not available for all children. Consequently, as explained further below, other different strategies are used to correct for the measurement error bias.

Another option to face the problem of correlation between nutritional status and the error component is by assuming to have a sample of siblings, with available information on schooling and health. Assuming constant impacts between household members, the use of "within-siblings" approach has the advantage of purging any household and environment unobserved inputs that are constant across siblings[6].

Within-siblings device is not enough for the purpose of this paper. As a matter of fact, it is still possible to have a correlation between the child's health status and the error term. This correlation can be removed with the addition of instrumental variables (IV) to this estimation. An appropriate instrument has to satisfy two main conditions: those of relevance and of exogeneity. The first means

[6] Within-sibling data long have been used in the social sciences, since early works by Chamberlain and Griliches, 1975; Behrman and Taubman, 1976; Olneck, 1977. However the within-sibling approach is limited in several ways, since, for instance, it will only remove across family differences, but not within family differences.

that an instrument must explain a substantial proportion of variation in the endogenous variable. The second implies that the instrument must only have an effect on the outcome of interest through the endogenous variable and cannot be correlated with the residual. Hence, in order to be a good instrument, incidence of transitory shocks should satisfy both of these conditions. Bearing in mind, the aim of this analysis, it is very hard to find such a variable since most factors influencing the decision of investing in health will also influence the decision of investing in education.

The model assumes that the child's health status in a given period is partly due to a health shock or a combination of shocks (which work as a "natural" experiment) that takes place after parents have made decisions for the same time period. This kind of shock would be (i) of adequate magnitude and persistence to affect a child's health variable used, (ii) adequately variable across siblings in the same household, (iii) adequately transitory not to affect the sibling's health variable used and, (iv) not correlated with educational performance.

To resume, since the naïve estimates assume that nutrition is exogenous (pre-determinate) and lead to biased estimates of the parameters, the preferred estimation strategy here adopted is a household fixed effects (FE) with IV approach. In presenting the estimation strategy, the choice of this approach will be confirmed by additional reasons.

3. Description of data

3.1. Kagera Health and Development Survey (KHDS)

This study uses baseline data from a longitudinal Living Standards Measurement Survey (LSMS) conducted in Kagera, a rural region located in North West Tanzania.[7] The study was carried out by the World Bank, Muhimbili University College of Health Sciences and University of Dar es Salaam. The Kagera Health and Development Survey (KHDS) is a longitudinal socio-economic detailed survey, into the long-run wealth dynamics of households and individuals. The data holds a rich set of

[7] Since the present work uses baseline data from a region of Tanzania, the author analysed the country profile in order to have a proper view of the topic investigated. For those interested in a closer examination of this part, please contact the author. However, here is interesting to point out that the economy of Tanzania is predominantly agrarian, with agriculture providing over 80% of all employment. Poverty remains overwhelming in rural areas where about 87% of the poor population live, performing low rates of health and education.

community characteristics on health care, education and prices as well as individual and household specific information.

The Kagera region is representative of many parts of sub-Saharan Africa; therefore, its study is important as it may reveal many important aspects similar to those countries. The region takes its name from the Kagera River; it lies just south of the equator, on the western shore of Lake Victoria, bordering Uganda to the north and Rwanda and Burundi to the west.

The survey is a four-round panel from 1991 to 1994 followed up in 2004.[8] The KHDS used a random sample that was stratified geographically and according to several measures of adult mortality risk. The first KHDS consisted of 915 households effectively interviewed up to four times, from September 1991 to January 1994. Households were drawn from 51 villages (or "clusters") of 16 households each in the 6 administrative districts of Kagera[9]: Biharamulo, Bukoba Rural, Bukoba Urban, Karagwe, Muleba and Ngara.

In 2004, a fifth round of data collection was completed. The objective of the KHDS 2004 survey was to re-interview all individuals who were household members in any round of the KHDS 1991-1994 and who were alive at the previous interview. The KHDS 2004 intended to provide data to revise economic mobility and variations in living standards for the sample of individuals interviewed in the decade before (Beegle et al., 2006). Considerable effort was made to track surviving respondents to their current location, be it in the same or a nearby community, within the region, or even outside the region. Excluding households in which all previous members were deceased, the field team managed to re-contact 93% of the baseline households in 2004. This is an excellent rate of re-contact compared to panel surveys in higher income countries.

[8] For a detailed description and data set of the first four waves see World Bank (2004) and Living Standards Measurement Study website. For further information see Ainsworth et al. (1992) and World Bank (1993b). Concerning the follow-up survey see Beegle, de Weerdt, Dercon (2006).

[9] During the first passage, a total of 840 households were interviewed. This group included the 816 "original" households selected from the enumeration (or their replacements) and 24 "extra" households. By the end of the fourth passage, 81 households (9.6% of the 840 interviewed in the first passage) had dropped out and this attrition rate compares favourably to other panel data sets (Alderman et al., 2001). The main cause for such little rate of shrinkage in size was that the household moved, principally due to an adult death in the household.

3.2. Rainfall pattern and data

Development in Tanzania has been vulnerable and adversely affected by inter annual variations in both weather and climate and their related disasters, such as drought and floods (caused by excess or deficit of rainfall respectively). Drought is characterized in terms of its spatial extension, intensity and duration and a precise universal definition for it is not easy to find.

In recent years, parts of Tanzania have experienced recurring droughts. The most devastating were those of 1983-1984 and 1993-1994. According to Tanzanian historical data, droughts occur every four years, affecting over 3.63 million people. Differently from the general pattern, data of KHDS elaborated for the paper show that the period of severe droughts occurred between 1988 and 1990.

Kagera region presents a bimodal seasonal calendar in a 12 month with 900-2,000 mm per annum, temperatures range between 20°C and 28°C, and the timing and duration of each season vary according to setting within the region. It presents long rains (the masika season) from March to May and short rains (the vuli season) from October to December; dry season (the kiangazi season) between the two rainy seasons.

To perform the declared analysis, this paper uses monthly rainfall data collected by Tanzania Meteorological Agency and reported on the EDI (Economic Development Initiatives) website. Historical rainfall data are available for 25 years, from 1980 to 2004; total millimetres of rain per month and total days of rain per month for 21 weather stations in Kagera region were collected.

Average annual rainfall in Tanzania varies considerably between areas. Kagera is not the driest region throughout Tanzania, given its location; however, its weather is still of considerable impact on people's well-being. For agricultural production, however, it is not the average annual rainfall which matters the most, but its seasonal pattern[10]. The increased vulnerability of households to weather shocks and their sharp dependence on the timing and quantity of rainfall is emphasized by the 2008 WDR (World Bank, 2008)[11].

---

[10] Corresponding to the bimodal seasonal pattern, the agricultural season starts in October. The distribution of rainfall during the short rainy season is critical, as it determines when fields preparations can start and the success of early crops. The second critical period in relation to rainfall is the length of the dry spell between the short and long rains.

[11] The only analysis based on a part of rainfall data (ten years historical trend) assembled by EDI website are two. The first by Litchfield and McGregor (2008), which connect rainfall in 2004 and welfare in Kagera,

The season during which a child is born is one crucial factor for his nutrition and health status. The impact of season of birth becomes less significant after the third year of a child's life. The rainfall occurring the season before the child is born is extremely important, given that it probably affects production more and, therefore, the availability of food for a child's nutrition. The season and period during which a child is delivered is also crucial. During both long or short rainy seasons, there usually is a higher agricultural production and this requires higher (also feminine) labour force employed in the fields. For a child born during these rainy seasons, the consequences are twofold: less care by mothers who are usually the main care giver[12], with negative effects on child's health; higher family wealth, which positively affects both the child's and the other existent family members' health.

Moreover, higher early-life rainfall is supposed to increase spread of diseases such as malaria, which in turn negatively affects infant health status[13]. As revealed by data elaboration, malaria is the first cause of death in Kagera and prevalent in Tanzania. The risk to contract malaria (principally as P. falciparum) is present throughout the calendar year in Tanzania, under 1800m of altitude[14].

4. Estimation approach and sample investigation

4.1 Estimation approach

The estimating strategy is based on household FE-IV estimator. Differencing among siblings purges any household and environment inputs, both observed and unobserved, that are constant

showing that the weather shocks have larger effect for the poorer households than for the wealthier. The second one by Trærup and Mertz (2009), which explores the nature of shocks that rural households in Kagera have experienced, examines their responses to these shocks and assesses the impact of future climatic changes on these responses.

[12] Artadi (2006) shows that in sub-Saharan Africa there is a trade-off between infant health and family income and that households living in areas where such trade-off is present, tend to choose the worse birth months for infant survival. The author recognises that Tanzania has a strong seasonal pattern of malaria, and higher precipitation around the time of birth is associated with higher infant mortality. Moreover, she reveals that the rural population in the country does not face such trade-off, and that concentrates births when survival is higher.

[13] According to the World Malaria Report 2005 (WHO and UNICEF, 2005), Tanzania is classified as a malaria endemic country, subject to frequent and recurrent epidemics, mainly generated by anomalies of rainfall and/or temperature. 93% of the population is living in areas at risk of stable malaria. Getting precise information about incidence and trends of malaria in Tanzania is not easy, mainly because malaria is not considered a notifiable disease in the country and because people, especially in rural areas, tend to cure diseases at home without seeking care to formal health providers. Concerning Kagera, information about the incidence of malaria is not easily available, but the region is classified as being at high risk of malaria and easily subject to serious outbreaks.

[14] Given this consideration, a dummy variable is elaborated regarding possibility of contracting malaria, in relation to the altitude of the nearest station.

among them. Rainfall can be used to overcome the endogeneity and simultaneity problems of health measure. As argued above, the weather shock has to be the one that affect H_{i1}, varies across siblings in the same household and be sufficiently temporary not to affect S_{i2}. Weather shock is conceivably related to difference in sibling's height-for-age, but plausibly, no persistent effects on schooling outcomes are observed[15]. While there might not be any existing evidence that rainfall considered in this study have any effect on the achievement production process, it remains still possible.

To ensure the validity of the indicated strategy, it has to be considered that environmental or social factors as, in this case, a weather shock occurred during the first years of life, can eventually be correlated not only with nutrition and health, but might also affect cognitive ability throughout different channels. For this reason, a correct analysis of the composition of the error term is crucial to understand whether the rainfall instrument is acceptable.

A severe drought - or other kinds of shock - might alter the amount or the quality of child care provided by parents since they must spend more time in coping with the drought and in procuring enough food for the family, with probable negative consequences on education (Neumark, 1999; Ashenfelter and Kruger, 1992). Unusual rainfall might alter the disease environment, with direct effects on the absorption of nutrition particularly in the vulnerable early years of rapid growth. Fluctuations in precipitation may influence other environmental conditions correlated with economic activity and public health, such as the availability of potable water and agricultural pest control. Indeed, some of these channels may imply a negative impact of rainfall that would somewhat offset positive effects via improved agricultural output. Other kinds of shock could in turn create new sources of stimuli, with positive effects on schooling.

Differences between siblings were due to differences in their inherited talents and motivations, experiences and stimuli they were exposed to. The scope to which environment varies between siblings depends on their age difference and on the changing conditions of the household and community associated with this lapse of time (Griliches, 1979). Since the Kageran society is almost "static", as it is overwhelmingly rural and traditional, and since children included in the main sample

[15] Certainly the shock not only influence health status of children, but more probably it affects the economy as a whole. The approach used takes in consideration the differential impact of the selected shock on siblings; hence, the time it took the local economy to recuperate is not a feature in the analysis.

have a small age difference between them, it can be assumed that all siblings would be affected on the same measure by the shock. Robustness checks were offered on children with a reduced age difference. Given these considerations, the correlation between weather shock and cognitive ability would justify the inclusion of household FE, which would eliminate the ability differences between siblings that can be hardly observed by the researcher[16].

The academic input on which the analysis is focused would be the child's nutritional history during the early years of life. However, as a realistic issue, quantifying cumulative nutrition inputs is exceptionally difficult. A straightforward option for the nutritional history up to that age can be to use the child's anthropometric measure of height-for-age.[17] For comparison purposes, children's height measurement is standardized according to the International Referenced Population defined by the U.S. National Centre for Health Statistics (NCHS) with the Centres for Disease Control (CDC) and the World Health Organization (WHO) (WHO, 1995). High-for-age indicator is thus expressed as a "z-scores", which compares a child's measurements and gender with those of a similar child in a reference, healthy population defined by the US NCHS, who has a z-score with mean zero and standard deviation (SD) of one.

A low height-for-age z-score defines "stunting", which means the child is too short for the age. Stunting indicates slow physical growth since birth, usually due to repeated episodes of poor nutrition and/or episodes of diarrhoea and other illnesses. It is a cumulative indicator of past episodes of malnutrition and, since it is likely to persist even after these conditions are eliminated, it is also defined as indicator of chronic malnutrition.

A possible criticism of this measure can be that height-for-age also captures the consequences of illness and other environmental and genetic influences, reflecting more than just a child's cumulative nutritional history. For this reason, controls with the Body Mass Index (BMI) have been done at first analysis to corroborate the results obtained.

---

[16] Ashenfelter and Kruger (1992) surprisingly find no evidence on positive effect of unobserved ability to the final grade achieved; alternatively, they find some weak evidence on negative correlation of unobserved ability to schooling level.

[17] The WHO recommends stunting as a reliable measure of overall social deprivation (WHO, 1986). Several studies found it to be the most strongly associated with school progress compared to the other anthropometric measures of malnutrition (see for example Jamison, 1986; Moock and Leslie, 1986).

The respondent variable corresponding to educational achievement indicates whether or not the adolescent has completed the entire cycle of primary education. This choice is determined by some reasons. At first, given the data structure and the periods of time available, this one results the foremost information to analyze. For instance, primary school in Tanzania is mandatory (and free since the end of 2001); therefore, all children are supposed to attend and complete primary education. Nevertheless, as it is shown in the following paragraphs, these considerations do not reflect the real situation. Additionally, a child can have repeated a school year many times, but only if he has completed the final grade of primary school denotes the basic information required to pass the final exam have been adequately achieved. Consequently, the selected dependent variable, for example, health status, is not only a good indicator of individual characteristics, but it also reveals essential socio-economic elements, like parental preferences or cultural traditions.

A suitable schooling outcome can also be represented by the total years of schooling completed. The literature (see among the other works, those of Grossman, 1972; Fuchs, 1965) suggests that among socio-economic variables, years of formal schooling completed is probably the most important correlator of good health. This study is focused on a sample of children, the majority of whom are still attending school. Therefore, years of education attained do not necessarily indicate the total years to be achieved by the respondent if the child is still going to school and expected to complete additional years of education. Data concerning actual years of completed schooling are deeply analyzed and summarized, in order to avoid discarding of essential information. For the sake of clarity, the main regressions are analyzed, maintaining as respondent variables, both final grade achieved and the completion of primary school respectively; the results corroborate the choice of the second variable[18].

Besides, another important measure to be included in the analysis is the delay in primary school enrolment. Many children enter school after the minimum age at which they are allowed to enrol and this fact reveals parental assessment of whether their child is ready to go to school. This phenomenon

[18] The weird result obtained in regressing the final grade achieved on height-for-age is that the health status seems to negatively affect education. The same sign of relation appears between rainfall and education. For more details on those regressions, please ask the author.

has been noted in many low income countries, despite the human capital theory predicts that schooling will begin at the earliest possible age (see Glewwe and Jacoby 1995).[19]

As shown by the evidence presented in the following paragraphs, higher rainfall should be interpreted as a positive shock to Kageran population, leading to higher products from agriculture. From the analysis of the historical data on rainfall for a period of 25 years, computing the mean and SD for such long period of rainfall, the paper identifies whether rainfall at location and time of birth was substantially above or below normal levels, and therefore measures the effect of weather shocks on child’s human capital attainments. Specifically, exposure to transitory weather shocks experienced by children during their birth year can be assessed to identify differences in height-for-age of siblings as pre-schooler, and then connected to schooling outcomes during early adulthood.

4.2 Description of main variables

The purpose of the paper is to estimate equation (1), in order to analyse the long-term consequences of infant and early childhood health status on education outcomes. Given the estimation problems presented above, data requirements are high.

A serious problem in dealing with data set collected for developing countries is that people frequently do not have precise information concerning themselves. The suitable sample for the analysis is made up of children with available information on: health status in the first time period, educational outcomes in the second time period, the identified shock used to instrument health as described above, birth date and location of children to accurately associate and measure the effect of the shock to each child. The sample has to preserve the same children traced both in the first and in the second time period, without attrition bias between the two. In the selected model, in the first time period the individual is a newborn or a preschooler, and is interviewed for the first time during 1991-1994; in 2004, the second time period, the individual is an adolescent or a young adult. A complete list of all variables included in the estimation is presented in Table 1.

[19] A supposed negative correlation between delaying entrance and grade achieved exists. As a matter of fact, both final grade achieved and delay in enrolment are the observed outcome of the latent propensity for schooling. Often, information on enrolment and attainment differs because a child enrolled at the beginning of the year does not attend school later on.

Specifically, the analysis focuses on children born between March 1985 and January 1994, and traced in 2004. Data show that of 1,548 preschoolers in 1991-1994, some 1,214 were traced in 2004, of whom almost half are female. Of these children still in the sample in 2004, some observations have to be dropped because of missing data in fundamental variables.[20] Consequently, the applicable sample consists of 238 households with more than one child of preschool age in the first time period, for an amount of 622 not only-children (334 male and 288 female) (Table 2). All children are households members and have at some point attended school by 2004.

Table 3 and Table 4 provide some summary statistics. Table 4 shows that the mean height-for-age z-score is -1.65; almost 70% are moderately stunted, performing an height-for-age z-score under -1 SD below the mean of the international reference population; nearly 40% are severely stunted, with an height-for-age z-score under -2 SD[21]. These considerations are corroborated by both the histogram in Figure 1 and the bar graph in Figure 2, indicating an high concentration at the level of –2 SD. The condition of high chronic malnutrition among Kageran children reflects the general situation in Tanzania (see Appendix I).

A further consideration is that gender and residency are two important discriminating factors for possible malnutrition, as male and rural children are more likely to be stunted than female and urban children.

Several factors related to children re-interviewed in the second time period had to be considered. Here, only the most relevant factors are reported.[22] Even though primary school is compulsory from the age of seven in Tanzania, of the children in the selected sample, almost 22% enter school at the due time, while two third is enrolled between 8 and 10 years old. Some of them are still enrolled in

[20] In detail: 84 children lack information about birth month; 118 children lack information on education (principally the final grade achieved); 157 children lack information on anthropometric measures (height) and 7 children whose anthropometric measure has a plausible measurement error (height-for-age z-score is less that -6 or more than +6).
Concerning education, the first seven grade of primary education are counted with points from 1 to 7; the first year of secondary school is indicated with grade 8, with and ascending order for the following grades. If the child have some education indicated as “Adulted”, “Koranic” or “None”, all of them have been assigned education level equal to 0.

[21] These results are consistent with those collected by Alderman et al. (2005) and by TDHS data sets.

[22] Further descriptive statistics have been analysed by the author. Data show that children resident in rural areas are much more numerous than those in urban areas. In the former group, on average, children are shorter, achieve less grade of schooling and start school later than those in the second one. Additionally, males are slightly more numerous than females and very few little difference exist among them; this result reflects the socialist culture of Tanzania, impressed in both the health and educational system.

primary education, and therefore they have not terminated the primary school cycle. These results reflect the general situation of Tanzania.

Data concerning actual year of completed schooling are deeply analyzed and summarized, in order to avoid discarding essential information. An important observation from the analysis of the sample is that although children are almost equally distributed by age, besides the peak at the final year of primary school (grade 7), a second peak exists in the third year of primary school in Kagera. This fact can be determined by parental choices: plausibly, rural households need child's help in the fields or have limitation in financial resources that lead the child to drop out school after having achieved the rudimental knowledge of reading, writing and counting. Otherwise, it can be a consequence of some informal customs of the community. Another explanation can be that an institutional law in Tanzania, for example, provide for a blockage examination, which a child needs to succeed to proceed the schooling path[23]. Moreover, the abolition of fees for primary education in Tanzania could have an impact on this trend, and on the investments decision concerning all children both already and still not enrolled. Unlikely, the data set does not cover such a long period to examine long term consequences of the law on educational choice, and, therefore, on the second peak at grade 3. This point needs to be analyzed more in detail collecting new information on this population, because it probably hides some interesting information that could enrich the results of the research in general.

4.3. Potential selection biases

The sample employed for this study can be subject to some sources of attrition biases. The primary complexity in managing data set collected for developing countries is that people often do not have accurate information concerning central aspects of their life. The data set exploited in this paper does not make an exception. Consequently, a potential attrition bias may be determined by the missing data in fundamental variables for some children, such as birth date, the final grade achieved, starting school age and anthropometric data in one or both time periods. Given the scarcity of relevant

[23] The first researches on this topic did not reveal the existence of such kind of law, but there is still the chance of an informal social custom that drives to the same result. A similar trend has been revealed by studies in other countries, as the one of Maluccio et al. (2006) for Guatemalan, but no mention about this fact appears.

information for some children, a number of observations have to be dropped inevitably, with the consequent constraint of the sample.

Additionally, controlling for siblings of a similar age considerably narrows the sample, generating a possible selection bias.

Finally, the sample selection can be biased because of deaths occurred through the panel composition. More precisely, 131 children of preschool age (56% female), died between the two time periods. For the 90% of this group the principal cause of death was illness: the first cause was malaria by 23%; the second cause was associated to AIDS/HIV by 20%. Moreover, almost 90% was malnourished. Given these considerations, excluding from the sample those children who died essentially because they were particularly unhealthy, leads to upward biased estimates, unavoidably based only on survivors. Table 5 provides the same voices as Table 3 for children who died between the two time periods. As confirmed by the table, children who were dead through the panel composition compared to survivors (who built the original sample employed in the analysis) generally present worst levels of health. This consideration is driven by the assessment of height-for-age z-score under -2 SD.

5. Empirical results

5.1. Testing the validity of the instruments and estimates

The respondent variable corresponding to educational achievement indicates if the adolescent has completed or not the entire cycle of primary education. The explanatory variables for the individual i are: height-for-age, which represents the measure of stunting expressed in z-scores at the time of the first interview, occurred within 1991-1994; the age as adolescent in the second time period, and gender, which is time invariant. The estimating strategy is based on household FE-IV estimator using a linear probability model. Differencing among siblings purges any household and environment inputs, both observed and unobserved, that are constant among them.

Table 6 shows the results of estimating the impact of height-for-age as pre-schooler on educational outcomes as adolescent. The estimation approaches employed are four: the naïve Ordinary Least Squares estimate (OLS) with controls for time invariant child’s characteristics, like age as adolescent and gender; the household fixed effects estimate (FE); the instrumental variable estimate (IV) with district fixed effects; the household fixed effects - instrumental variables estimate (FE-IV), by means of the weather shock variable already described as the instrument.

All the independent variables have positive impact on the dependent variable. Females have slightly more chance to complete primary education, but it is not statistically significant. Age is statistically significant at the 1% level on education and its positive impact remains almost the same for all the estimation approaches. Certainly, the older the child, the higher the probability of having completed the primary school. Better pre-school health status is associated with greater educational attainments in all the estimation approaches. The effects of both un-instrumented and the instrumented approaches are roughly comparable in magnitude by couples. The un-instrumented estimates have almost halved marginal effects (but also lower robust standard errors) compared to those instrumented, remain still imprecise. Height-for-age variable appears statistically significant at the 1% level when using the un-instrumented estimates, while it is no more significant for those instrumented; this fact can be a result of the partial endogeneity of health for education. R^2, as a statistical measure of how well a regression line approximates real data points, is high for all the estimation strategies.

Table 7 provides the results of the first stage within siblings regression used to instrument the endogenous variable of child's height-for-age. Rainfall has a positive and statistically significant impact at the 1% level on initial height-for-age. As the summary statistics reveal (Table 1), rainfall value at location and time of birth expressed in z-score is globally included in the interval [-.336, 1.026]. In sub-Saharan African countries, rainfalls are very often the main source of safe water for drinking, personal hygiene and farming – which means nutrition for both the child and the care giver. Therefore, the more rainwater available (within safe bounds), the higher the probability of a child being healthier. Both gender and age as adolescent are statistically significant at the 1% level on initial height-for-age. While being a female has a positive impact on the early health status, the effect of age is negative, since the older the child, the smaller the gap with the reference healthy child.

Staiger and Stock (1997) employ as rule of thumb that the F-test statistic for joint significance of the instruments in the first-stage should exceed 10. The F-test statistic is particularly high and equal to

17.67 and 21.61 for IV and household FE-IV respectively. It can be concluded that the instrument is valid and not weak.

Concerning the reduced form estimates (not shown here), both gender and age as adolescent have positive effects on education. Being a female has a small positive impact on the option of completing primary education (gender has a minor effect on education compared to that on health as shown above), and it is statistically significant only when using the naïve estimate. Age is statistically significant at the 1% level on education and its positive impact remains almost the same for the two estimation approaches. Rainfall has a small positive, but not statistically significant impact on education. As already stated, the chosen instrument is not weak, although not much correlated with education.

Table 8 illustrates the results obtained with the addition of some controls to the estimates of the impact of height-for-age as pre-schooler on educational outcomes as adolescent. The independent variables as gender and age have more mitigated impacts on the dependent variable, compared to those presented in Table 6. Gender in household FE estimates here results negatively correlate with education, but the magnitude of the impact remains very small and not statistically significant. Age is statistically significant at the 1% level on education and its positive impact remains almost the same for all the estimation approaches, even slightly lower than in the regressions without controls. Concerning pre-school health status, estimates are almost similar to those presented in Table 6. The un-instrumented estimates have almost the same marginal effects and robust standard errors, while the instrumented estimates have reduced marginal effects compared to those of the previous estimations. Height-for-age variable appears statistically significant at the 1% level when using the un-instrumented estimates, while it is no more significant for those instrumented, corroborating the previous results. Household characteristics have relatively small, and in some cases very small, impact on educational achievements. At community level, all districts have negative statistically significant impact on education of almost the same magnitude. Living in urban or healthy areas, on the contrary, significantly increases the opportunity to be educated and to conclude the primary school. Several other variables were also included in earlier models, but were dropped due to significantly high correlation with other variables or to very few observations in some categories.

5.2. Estimates for a particular sub-sample of districts

In performing some checks, an interesting point arose. When each district is removed in turn, maintaining the variability by districts, the outstanding outcome is that height-for-age is significant in some cases (Table 9). More precisely, when district n.1 (Karagwe) or district n.5 (Ngara) is removed in turn, height-for-age is statistically significant at 10% level; additionally, eliminating both districts simultaneously the variable becomes more significant, with an increased magnitude.

Further investigation on these two districts has been carried out in order to find plausible reasons for such a different result.[24] When comparing the two problematic districts with the other four, what emerges is that although, all of them have similar values of education outcomes, gender and age, the former present differences concerning health status and weather conditions. More precisely, except for Biharamulu, children residing in Karagwe and Ngara districts have the worst health performance on an average.

Additionally, Karagwe and Ngara districts have negative average values of rainfall in z-score, meaning that the two areas are dryer in comparison with the other districts. In fact, these are the only two districts located far from Lake Victoria. Except for small lakes spread through the region, the two districts are characterized by more continental and thus, stable weather, since they are less subject to the influence of the lake. Rainfall in these two districts have a negative and not statistically significant value at the first stage regressions. This fact explains the reason why it is not significant when all districts are taken.

Apart from those considerations, the strongest explanation for the indicated difference among districts is the following. Since the early Nineties, extremist militia groups carried out the extermination of élite Tutsis and moderate Hutus in the genocides of Burundi (1993) and Rwanda (1994). Almost one million people were killed and thousands were compulsorily displaced from their homes. Over the course of a few months, Karagwe and Ngara districts of Kagera were the primary asylums for some 600,000 refugees from Burundi and Rwanda to escape ethnic violence in their home countries. Consequently, in Karagwe and Ngara districts, the road infrastructure was damaged through

[24] More information concerning descriptive statistics of the main variables of interest by districts are available. Please, if interested, ask the author.

over use, school and health facilities were overloaded, trees were cut extensively to make way to refugee settlements, some diseases spread[25]. The damages wrought by the refugees in the two districts of Kagera after half a year were estimated to require at least $65 million to be repaired (Smith, 1995). Despite these serious problems, the number of refugees continued to grow till double the locals in number over the time.

Table 10 provides some summary statistics related to the sub-sample of children, obtained excluding those living in districts n.1 and n.5. Exception for all the other variables, which have essentially the same values of those of the original sample examined in Table 4, the height-for-age z-score for pre-schooler in this sub-sample of children is -1.5.

Table 11 shows the results of estimating the impact of height-for-age as pre-schooler on educational outcomes as adolescent for the sub-sample of children (the regressions are the same as for the original sample). The independent variables have positive (or null) impact on the dependent variable. The couple of un-instrumented and instrumented estimates have almost the same marginal effects and robust standard errors among them. The main difference with results obtained in Table 6 is that, here, height-for-age variable appears statistically significant not only when using the un-instrumented, but also with the instrumented estimates.

The first stage estimates report almost the same values as those of the original sample, being slightly greater in magnitude (Table 12). Similar considerations are worth concerning the reduced form estimates, not shown here. Rainfall has doubled its effects on education, becoming now statistically significant at the 1% level.

Controls are added to the estimates of the impact of height-for-age as pre-schooler on educational outcomes as adolescent in the four selected districts (Table 13). Being a female has now a still small but negative impact on the dependent variable, being statistically significant only with the household FE–IV estimation approach. Age is statistically significant at the 1% level on education and its positive impact remains almost the same as that in the original sample, being slightly higher when

[25] Baez (2008) investigates the causal effects of hosting Burundi and Rwanda's refugees on the outcomes of Kageran children. The author finds evidence of adverse impacts on children's anthropometrics, the incidence of infectious diseases and the mortality for children under five almost 1.5 years after the shock. He also finds that childhood exposure to this massive arrival of refugees reduced height in early adulthood by 1.8 cm (1.2%), schooling by 0.2 years (7.1%) and literacy by 7 percentage points (8.6%).

applying the instrumented approaches. Pertaining to the pre-school health status, it appears statistically significant when using the household FE and all the estimates, except for OLS, present higher or even much higher values than those of the original sample. The value obtained by applying the household FE-IV approach is high and statistically significant for this sub-sample, resulting in an outstanding result. At community level, differently than the complete sample, all districts have positive statistically significant impact on education of almost the same magnitude among them. Living in urban areas increases the opportunity to be educated and to conclude the primary school, and the magnitude is doubled compared to the original sample.

Examining the results relating to the sub-sample, the level of the impact of health status on educational outcome is meaningful. The height-for-age z-score for pre-schooler is -1.5 as already indicated in Table 10 (or eventually -1.65 for the complete sample). If the population had the nutritional status of well-nourished children of the reference population, the mean z-score would be 0. Applying the household FE-IV parameter estimates reported in Table 13, this would result in an additional 28% probability of completing primary education.

5.3. Robustness checks

This section presents robustness checks to corroborate the results illustrated above and to find possible explanations regarding the missed causal link between pre-school anthropometric status and subsequent educational achievements.

The scope to which environment varies between siblings depends on their age differences and on the changing conditions of the households and the communities associated with this lapse of time (Griliches, 1979). The correlation between weather shock and cognitive ability has been used to justify the inclusion of household FE, in order to eliminate the ability differences between siblings. To strengthen the choice of such expedient, assuming that all siblings would be affected on the same measure by the shock, the sample can be restricted to children who have a difference of age minor to three years. Even though the sub-sample is remarkably reduced in size, Table 14 (a) shows almost the same results compared to those of the original sample (Table 6). The same is worth pertaining to height-for-age variable, which is statistically significant at the 1% level when using the naïve estimates, while it is no more significant for the household FE-IV. Equivalent result of height-for-age on education can be obtained applying a threshold of age difference of two years within children of the same household (Table 14 (b)). The sub-sample become very small, thus the values cannot be considered precise representation of the original sample.

Similarly, few considerations can be driven regarding regressions on siblings with restricted age difference for the sub-sample with the selected four districts. Although the sub-sample is reduced in size, Table 15 (a) shows the same important outcome pertaining to the height-for-age variable, meaning that it is statistically significant at the 5% level even for the household FE-IV. Applying a threshold of age difference of two years within children of the same household (Table 15 (b)), the new sub-sample becomes very tiny. Not much can be said, except for the positive sign of height-for-age in the household FE-IV, in comparison with Table 14 (b).

6. Final considerations

In recent years the crucial role of human capital in economic growth and poverty alleviation has been almost unanimously stated. The correlation between mental and physical preschooler health and education outcomes has been widely recognized. Nevertheless, the exact mechanisms by which these outcomes are related is only progressively being revealed and remains a considerable challenge.

Extending the literature on the determinants of human capital formation, the central results reported in this paper are based on household FE-IV estimates. In order to drive an innovative kind of study, this study takes into account the endogeneity bias by instrumenting it and using weather shocks at the community level. External shocks may cause permanent damage to children's future welfare and cognitive abilities; therefore, further empirical investigation helping to quantify the magnitude of the effect of such shocks on early child growth is required.

Consistent with the vast literature, the principal finding is that an infant's and a child's health status play a key role in schooling outcomes. This result comes after having isolated the two districts on the western board, which had to cope with the difficult situation consequent to the settlement of refugees escaping from the genocides of Burundi and Rwanda in the early Nineties. Applying the household FE-IV approach, a child in good health during infancy has an additional 28% probability of completing primary education.

Almost all the literature on the impact of health on education, ignores the fact that individuals and households make choices in response to important predetermined characteristics unobservable for researchers and policymakers. The results confirmed in this paper has potentially important policy implications. The long-term effects of early-life conditions on schooling as adolescent should be factored into cost-benefit analyses of programs targeting this part of the population. Similar interventions are the promotion of exclusive breastfeeding, integrated child care and development programs and those providing nutritious supplements to pre-schoolers; all of these interventions have benefits and high rate of returns (Behrman et al., 2004). Future research on this important topic will undoubtedly benefit from the collection of high quality longitudinal data that recognizes the relationship between nutrition and human capital accumulation.

7. References

Ainsworth M., Koda G., Lwihula G., Mujinja P., Over M., Semali I. (1992), "Measuring the impact of fatal adult illness in Sub-Saharan Africa", Living Standards Measurement Study n. 90, World Bank.

Alderman H., Behrman J.R., Lavy V., Menon R. (2001), "Child Health and School Enrolment: A Longitudinal Analysis." *Journal of Human Resources* 36(1):185–205.

Alderman H., Hoddinott J., Kinsey B. (2006), "Long Term Consequences of Early Childhood Malnutrition," *Oxford Economic Papers*, 58(3), 450-474.

Alderman H., Hoogeveen H., Rossi M. (2005), "Reducing Child Malnutrition in Tanzania. Combined Effects of Income Growth and Program Interventions" World Bank, Policy Research Working Paper 3567.

Artadi E. V. (2006), "Going into Labor: Earnings vs. Infant Survival in Rural Africa", mimeo Harvard University.

Ashenfelter O., Krueger A. (1992), "Estimates of the Economic Return to Schooling from a New Sample of Twins," NBER Working Papers 4143, NBER, Inc.

Baez J.E. (2008), Civil Wars Beyond Their Borders: The Human Capital and Health Consequences of Hosting Refugees, IZA Discussion Paper No. 3468

Beegle K., de Weerdt J., Dercon S. (2006), Kagera Health and Development Survey 2004. Basic Information Document, World Bank, Washington DC.

Behrman J.R, Alderman H., Hoddinott J. (2004), "Hunger and Malnutrition". In Bjorn Lomborg (ed.) *Global Crises, Global Solutions.* Cambridge University Press

Behrman J.R. (1988) "Nutrition, Health, Birth Order and Seasonality: Intra-household Allocation in Rural India." Journal of Development Economics 28:1, no. February: 43-63.

Behrman J.R. (1996), The Impact of Health and Nutrition on Education. *World Bank Research Observer*, 11(1):23–37.

Behrman J.R., Deolalikar A.B. (1988), "Health and Nutrition." Handbook on Economic Development. editors Hollis B. Chenery, and T.N. Srinivasan, 631-711. Vol. 1. Amsterdam: North Holland Publishing Co., 1988.

Behrman J.R., Taubman P. (1976), "Intergenerational Transmission of Income and Wealth", *American Economic Review* 66:3 (May), 436-440.

Chamberlain G., Griliches Z. (1975), "More on Brothers", in Taubman P. (ed), *Kinometrics: Determinants of Socio-Economic Success Within and Between Families*. Amsterdam

EDI, Economic Development Initiatives (2004), Kagera Rural CWIQ. Baseline Survey on Poverty, Welfare and Services in Kagera Rural Districts.

FEWS NET - www.fews.net/tanzania

Fletcher J.M., Lehrer S.F. (2009), "Using Genetic Lotteries Within Families to Examine the Causal Impact of Poor Health on Academic Achievement", NBER Working Paper 15148

Glewwe P. (2005), The Impact of Child Health and Nutrition on Education in Developing Countries: Theory, Econometric Issues, and Recent Empirical Evidence, *Food and Nutrition Bulletin* 26(2):S235-S250.

Glewwe P., Jacoby H. (1995), "An Economic Analysis of Delayed Primary School Enrollment and Childhood Malnutrition in a Low Income Country." *Review of Economics and Statistics*, 77(1): 156–69.

Glewwe P., Jacoby H., King E. (2001), "Early Childhood Nutrition and Academic Achievement: A Longitudinal Analysis." Journal of Public Economics; 81(3):345–368.

Glewwe P., Miguel E. (2008), "The Impact of Child Health and Nutrition on Education in Less Developed Countries." *Handbook of Development Economics*, Vol.4.

Griliches Z. (1979), "Sibling Models and Data in Economics: Beginnings of a Survey," Journal of Political Economy, University of Chicago Press, vol. 87(5), pages S37-64.

Grossman M. (1972), "On the Concept of Health Capital and the Demand for Health", Journal of Political Economy, University of Chicago Press, vol. 80(2), pages 223-55.

Hoddinott J., Kinsey B. (2001), "Child Growth In The Time Of Drought." *Oxford Bulletin of Economics and Statistics*, 63, 409-436.

Lindeboom W., Kilama B. (2005), "Trends and Determinants of Malnutrition in Tanzania", Paper presented at the 10th Annual REPOA Research Workshop.

Litchfield J., McGregor T. (2008), "Poverty in Kagera, Tanzania: Characteristics, Causes and Constraints," PRUS Working Papers 42, Poverty Research Unit at Sussex, University of Sussex

Maluccio J.A., Hoddinott J., Behrman J.R., Martorell R., Quisumbing A.R., Stein A.D. (2006), "The Impact of Nutrition during Early Childhood on Education among Guatemalan Adults" PIER Working Paper 06-026

Moock P.R., Leslie J. (1986) "Childhood Malnutrition and Schooling in the Terai Region of Nepal." *Journal of Development Economics* 20(1):33-52.

Neumark D. (1999), "Biases in twin estimates of the return to schooling," Economics of Education Review, Elsevier, vol. 18(2), pages 143-148, April.

Olneck M.R. (1977) "On the Use of Sibling Data to Estimate the Effects of Family Background, Cognitive Skills, and Schooling: Results from the Kalamazoo Study" In Taubman P. (ed), *Kinometrics: Determinants of Socio-Economic Success within and between Families*. Amsterdam.

Pollitt, E. (1984), "Nutrition and Educational Achievement." Nutritional Education Series, Issue 9. ED-84/WS/66. UNESCO, Paris

REPOA (2009), "Institutional Analysis of Nutrition in Tanzania", mimeo

Rosenzweig M.R., Wolpin K.I. (1988), "Heterogeneity, Intra-family Distribution, and Child Health." *Journal of Human Resources* 23(4):437-461.

Schultz T. P. (1999), "Health and schooling investments in Africa", Journal of Economic Perspectives 13, 67—88.

Smith C.D. (1995), The Geopolitics of Rwandan Resettlement: Uganda and Tanzania, African Studies Association, *Issue: A Journal of Opinion*, Vol. 23, No. 2, Rwanda, pp. 54-57, http://www.jstor.org/stable/1166508

Staiger D., Stock J.H. (1997), " Instrumental Variables Regression with Weak Instruments" Econometrica 65, 557–586.

Strauss J., Thomas D. (1995), "Human Resources: Empirical Modelling of Household and Family Decisions." in *Handbook of Development Economics*, eds. J.R. Behrman, and T.N. Srinivasan, 1883-2024. Vol. 3. Amsterdam: North-Holland Publishing Company.

Strauss J., Thomas D. (1998), "Health, Nutrition and Economic Development." Journal of Economic Literature 36:2, n.June (1998): 766-817.

Trærup S., Mertz O. (2009), Linking climate trends to coping strategies in northern Tanzania, IOP Conference Series: Earth and Environmental Science, 6(41), pp. 412005.

TDHS 1991-1992, 1994, 1995, 1996, 1999, 2003-2004, Tanzania Demographic and Health Surveys, http://www.measuredhs.com/countries/country_main.cfm?ctry_id=39&c=Tanzania

WHO and UNICEF (2005), "World Malaria Report 2005", Geneva, 2005.

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

World Bank - Living Standards Measurement Study: http://www.worldbank.org/lsms/

World Bank (1993a), World Development Report 1993: Investing in Health. New York: Oxford University Press.

World Bank (1993b), "Report of a Workshop on: the Economic Impact of Fatal Illness in Sub-Saharan Africa", World Bank, Washington, D.C. and the University of Dar es Salaam.

World Bank (2004) "User's Guide to the Kagera Health and Development Survey Datasets." World Bank, Washington DC.

World Bank (2008), World Development Report 2008: Agriculture for Development. The International Bank for Reconstruction and Development. The World Bank.

World Health Organization (1986), Use and Interpretation of anthropometric indicators of nutritional status, Bulletin of the WHO, Geneva.

World Health Organization (1995) *Physical Status: The Use and Interpretation of Anthropometry*, WHO Technical Report, No.854, Geneva 3-466.

8. List of Tables

Table 1: Description of the variables used and descriptive statistics, KHDS (1991-2004).

Variables	**Definitions of Variables**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
	Individual variables					
Gender (female)	1= Child is female; 0=Child is male	622	.463	.499	0	1
Age as adolescent (in year) in 2004	Child's age as adolescent (in months) in 2004	622	1.473	2.381	10	20
Height-for-age z-score	Height-for-age for pre-schoolers (1991-1994) z-score statistics	622	-1.648	1.503	-6	4.08
Vaccine	1= Child has a Vaccination Card; 2=no	537	1.091	.288	1	2
	Household variables					
Small household	1= Child lives in a household with 2 children maximum, 0=elsewhere	622	.453	.498	0	1
Education of household head	1=Household head has completed primary education, 0=no	622	.436	.496	0	1
Father's height	Father's height (cm)	497	1.678	8.675	100	183
Mother's height	Mother's height (cm)	577	1.572	6.087	139	178
	Community variables					
Karagwe	1= Households located in Karagwe District, 0=no	622	.172	.378	0	1
Bukoba Rural	1= Households located in Bukoba Rural District, 0=no	622	.281	.450	0	1
Bukoba Urban	1= Households located in Bukoba Urban District, 0=no	622	.193	.395	0	1
Muleba	1= Households located in Muleba District, 0=no	622	.169	.375	0	1
Biharamulo	1= Households located in Biharamulo District, 0=no	622	.080	.272	0	1
Urban	1= Child lives in urban area, 0=no	622	.254	.436	0	1
Rainfall variation z-scores	Rainfall at location and time of birth z-score statistics	622	.036	.232	-.336	1.026
Household location (possibility of malaria)	1=Household located under 1800m of altitude (where malaria is more likely), 0=elsewhere	622	.254	.436	0	1

Source: Author's elaboration from KHDS data set

Table 2: Distribution of not-only-children in each household, by order of birth.

Birth order	Freq.	Percent
1	238	38.26
2	238	38.26
3	97	15.59
4	33	5.31
5	10	1.61
6	3	0.48
7	2	0.32
8	1	0.16
Total	**622**	**100.00**

Source: Author's elaboration from KHDS data set

Table 3: Heath status of children in KHDS 1991-1994, represented by "stunting" measure.

Variable	Gender		Residence		Total
	Female	Male	Urban	Rural	
Height-for-age z-score<-1 SD	64.93%	74.85%	66.46	71.55	70.26%
Height-for-age z-score<-2 SD	31.94%	47.31%	32.91	42.67	40.19%

Source: Author's elaboration from KHDS data set

Table 4: Descriptive statistics for children in KHDS 1991-1994.

Variable	Obs	Mean	Std. Dev.
Height-for-age z-score	622	-1.65	1.50
% Stunted	622	0.70	0.46
Age (in months)	622	32.58	24.87
% Female	622	0.46	0.50

Source: Author's elaboration from KHDS data set

Table 5: Heath status of children who dead between KHDS 1991-1994 and KHDS 2004, represented by "stunting" measure.

Variable	Gender		Residence		Total
	Female	Male	Urban	Rural	
Height-for-age z-score<-1 SD	72.41%	76.71%	65.52	77.45	74.81%
Height-for-age z-score<-2 SD	62.07%,	57.53%	48.28	62.75	59.54%

Source: Author's elaboration from KHDS data set

Table 6: Estimates of the education achievement equation for siblings.

Estimation Approach	OLS	FE (3)	IV	FE-IV (3)
Gender (female)	0.0377	0.0261	0.0243	0.0141
	(0.026)	(0.032)	(0.037)	(0.038)
Age in adolescence (in months)	0.115***	0.110***	0.120***	0.117***
	(0.0064)	(0.0080)	(0.010)	(0.013)
Height-for-age z-score	0.0471***	0.0379***	0.0838	0.0795
	(0.0082)	(0.011)	(0.066)	(0.061)
Constant	-1.380***	-1.310***	-1.383***	
	(0.080)	(0.11)	(0.084)	
Observations	622	622	622	622
Number of hh		238		238
R-squared	0.38	0.41	0.36	0.39

Notes: 1) Robust standard errors in parentheses. * Significant at 10%; ** Significant at 5%; *** Significant at 1%. 2) Standard errors for all the estimates are robust to clustered (village) sample design. 3) FE and FE-IV are estimated using a linear probability model.

Table 7: First-stage within siblings regression used to instrument endogenous variable of child's height-for-age.

Estimation Approach	IV	FE-IV (3)
Gender (female)	0.358***	0.278**
	(0.12)	(0.13)
Age in adolescence (in months)	-0.107***	-0.135***
	(0.027)	(0.031)
Rainfall in z-score	0.748***	1.187***
	(0.24)	(0.26)
Constant	-0.272	
	(0.41)	
Observations	622	622
Number of households		238
R-squared	0.07	0.13
F-test statistic	17.67	21.61

Notes: 1) Robust standard errors in parentheses. * Significant at 10%; ** Significant at 5%; *** Significant at 1%. 2) Standard errors for all the estimates are robust to clustered (village) sample design. 3) FE-IV are estimated using a linear probability model.

Table 8: Estimates of the achievement equation for siblings, controls included.

Estimation Approach	OLS	FE (3)	IV	FE-IV (3)
Gender (female)	0.0245	-0.0161	0.0264	-0.0224
	(0.036)	(0.044)	(0.047)	(0.056)
Age in adolescence (in months)	0.111***	0.107***	0.109***	0.113***
	(0.010)	(0.012)	(0.024)	(0.031)
Height-for-age z-score	0.0345**	0.0382**	0.0276	0.0639
	(0.015)	(0.017)	(0.11)	(0.12)
Vaccine	-0.0716	0.0247	-0.0740	0.0269
	(0.056)	(0.078)	(0.070)	(0.078)
Small household	0.0176		0.0171	
	(0.037)		(0.038)	
Education of household head	0.0930**		0.0947*	
	(0.045)		(0.050)	
Father's height	-0.00309		-0.00289	
	(0.0028)		(0.0041)	
Mother's height	0.00398		0.00422	
	(0.0029)		(0.0048)	
Karagwe	-0.219**		-0.222**	
	(0.098)		(0.098)	
Bukobar	-0.148**		-0.149**	
	(0.065)		(0.065)	
Bukobau	-0.354**		-0.361**	
	(0.14)		(0.17)	
Muleba	-0.204**		-0.204**	
	(0.090)		(0.087)	
Biharam	-0.220**		-0.223***	
	(0.084)		(0.083)	
Urban	0.220**		0.226*	
	(0.091)		(0.13)	
Household location (possibility of malaria)	0.200***		0.206**	
	(0.055)		(0.095)	
Constant	-1.445**	-1.275***	-1.510	
	(0.59)	(0.18)	(1.18)	
Observations	385	385	385	385
Number of hh		153		153
R-squared	0.34	0.31	0.34	0.31

Notes: 1) Robust standard errors in parentheses. * Significant at 10%; ** Significant at 5%; *** Significant at 1%. 2) Standard errors for all the estimates are robust to clustered (village) sample design. 3) FE and FE-IV are estimated using a linear probability model.

Table 9: Sub-samples of children removing one district in turn

	All districts but 1	All districts but 2	All districts but 3	All districts but 4	All districts but 5	All districts but 6	All districts but 1 & 5
Gender (female)	-0.0243	0.00398	0.0245	0.0314	0.0196	0.0274	-0.0212
	(0.046)	(0.044)	(0.047)	(0.039)	(0.039)	(0.041)	(0.047)
Age in adolescence (in months)	0.124***	0.123***	0.114***	0.116***	0.123***	0.100***	0.129***
	(0.013)	(0.017)	(0.018)	(0.013)	(0.012)	(0.014)	(0.012)
Height-for-age z-score	0.111*	0.118	0.0701	0.0373	0.0988*	0.0232	0.125**
	(0.062)	(0.081)	(0.10)	(0.065)	(0.051)	(0.063)	(0.051)
Observations	515	447	517	572	557	502	450
Number of hh	199	168	198	223	212	190	173
R-squared	0.37	0.33	0.38	0.43	0.37	0.38	0.36

Notes: 1) Robust standard errors in parentheses. * Significant at 10%; ** Significant at 5%; *** Significant at 1%. 2) Standard errors for all the estimates are robust to clustered (village) sample design. 3) FE-IV are estimated using a linear probability model. 4) District n.1 is Karagwe; district n.2 is Bukoba Rural; district n.3 is Muleba; district n.4 is Biharamulu; district n.5 is Ngara; district n.6 is Bukoba Urban.

Table 10: Descriptive statistics for children in KHDS 1991-1994, districts 1 and 5 removed.

Variable	Obs	Mean	Std. Dev.
Height-for-age z-score	450	-1.50	1.50
% Stunted	450	0.66	0.47
Age (in months)	450	32.63	25.07
% Female	450	0.46	0.50

Source: Author's elaboration from KHDS data set

Table 11: Estimates of the education achievement equation for siblings, districts 1 and 5 removed.

Estimation Approach	OLS	FE (3)	IV	FE-IV (3)
Gender (female)	0.0171	0.0146	-0.0194	-0.0212
	(0.031)	(0.037)	(0.044)	(0.047)
Age in adolescence (in months)	0.117***	0.116***	0.129***	0.129***
	(0.0072)	(0.0091)	(0.013)	(0.012)
Height-for-age z-score	0.0451***	0.0455***	0.130*	0.125**
	(0.011)	(0.012)	(0.069)	(0.051)
Constant	-1.403***	-1.376***	-1.436***	
	(0.091)	(0.13)	(0.11)	
Observations	450	450	450	450
Number of hh		173		173
R-squared	0.38	0.43	0.30	0.36

Notes: 1) Robust standard errors in parentheses. * Significant at 10%; ** Significant at 5%; *** Significant at 1%. 2) Standard errors for all the estimates are robust to clustered (village) sample design. 3) FE and FE-IV are estimated using a linear probability model.

Table 12: First-stage within siblings regression used to instrument endogenous variable of child's height-for-age, districts 1 and 5 removed.

Gender (female)	0.423***	0.452***
	(0.14)	(0.16)
Age in adolescence (in months)	-0.121***	-0.126***
	(0.030)	(0.034)
Rainfall in z-score	0.812***	1.572***
	(0.27)	(0.29)
Constant	0.0384	
	(0.46)	
Observations	450	450
Number of hh		173
R-squared	0.09	0.18

Notes: 1) Robust standard errors in parentheses. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 13: Estimates of the achievement equation for siblings, controls included, districts 1 and 5 removed.

Estimation Approach	OLS	FE (3)	IV	FE-IV (3)
Gender (female)	-0.0360	-0.0778	-0.0698	-0.152*
	(0.044)	(0.054)	(0.059)	(0.079)
Age in adolescence (in months)	0.109***	0.113***	0.127***	0.143***
	(0.013)	(0.016)	(0.027)	(0.030)
Height-for-age z-score	0.0271	0.0597***	0.111	0.189*
	(0.020)	(0.021)	(0.10)	(0.10)
Vaccine	-0.0434	0.0310	-0.0148	0.0932
	(0.082)	(0.10)	(0.095)	(0.12)
Small household	0.0318		0.0275	
	(0.046)		(0.052)	
Education of household head	0.0681		0.0378	
	(0.052)		(0.064)	
Father's height	-0.00782**		-0.0110**	
	(0.0034)		(0.0055)	
Mother's height	0.00210		-0.000446	
	(0.0040)		(0.0051)	
Bukobar	0.458***		0.387*	
	(0.16)		(0.23)	
Muleba	0.408***		0.336	
	(0.15)		(0.23)	
Biharam	0.384**		0.342	
	(0.16)		(0.22)	
Urban	0.496***		0.420*	
	(0.14)		(0.21)	
Household location (possibility of malaria)	0.192***		0.116	
	(0.071)		(0.11)	
Constant	-0.946	-1.318***	0.00676	
	(0.80)	(0.23)	(1.42)	
Observations	254	254	254	254
Number of hh		102		102
R-squared	0.34	0.33	0.28	0.18

Notes: 1) Robust standard errors in parentheses. * Significant at 10%; ** Significant at 5%; *** Significant at 1%. 2) Standard errors for all the estimates are robust to clustered (village) sample design. 3) FE and FE-IV are estimated using a linear probability model.

Table 14: Robustness checks. Sub-sample of children in household with difference of age <=3 years (a) and <=2 years (b), full sample.

Estimation Approach	OLS (a)	FE-IV (3) (a)	OLS (b)	FE-IV (3) (b)
Gender (female)	0.0237	0.0240	-0.00631	-0.0197
	(0.030)	(0.042)	(0.039)	(0.055)
Age in adolescence (in months)	0.121***	0.112***	0.123***	0.0978***
	(0.0089)	(0.027)	(0.013)	(0.033)
Height-for-age z-score	0.0436***	0.0678	0.0443***	-0.00266
	(0.010)	(0.059)	(0.014)	(0.10)
Constant	-1.451***		-1.457***	
	(0.11)		(0.17)	
Observations	386	386	228	228
Number of hh		183		109
R-squared	0.36	0.16	0.36	0.11

Notes: 1) Robust standard errors in parentheses. * Significant at 10%; ** Significant at 5%; *** Significant at 1%. 2) Standard errors for all the estimates are robust to clustered (village) sample design. 3) FE-IV are estimated using a linear probability model.

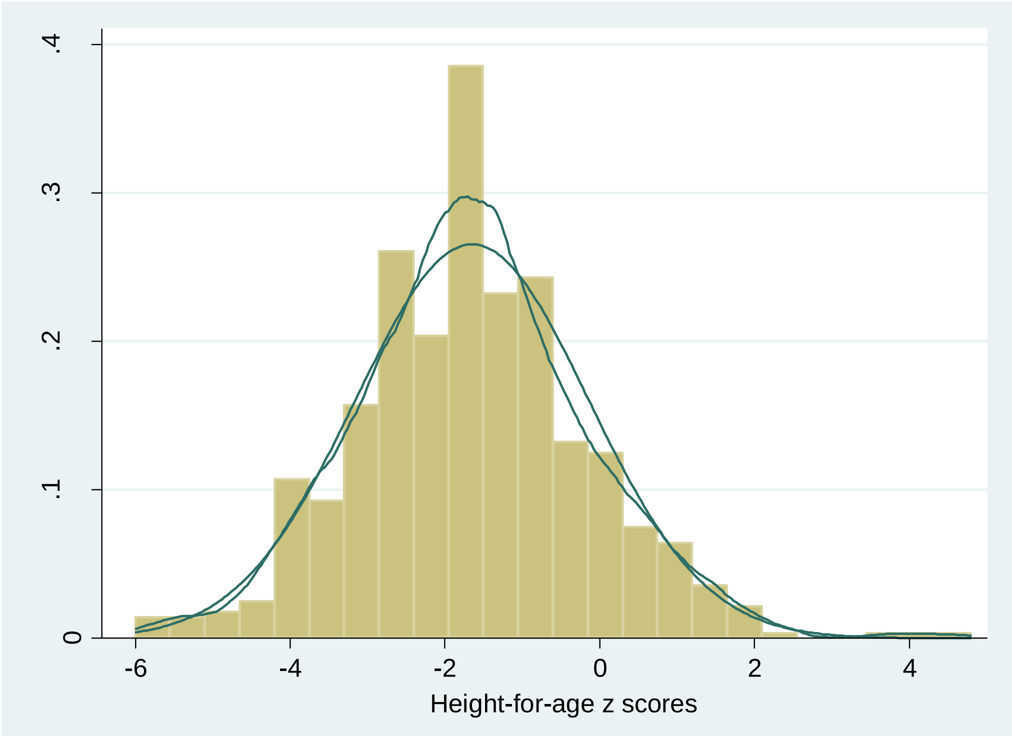
Table 15: Robustness checks. Sub-sample of children in household with difference of age <=3 years (a) and <=2 years (b), districts 1 and 5 removed.

Estimation Approach	OLS (a)	FE-IV (3) (a)	OLS (b)	FE-IV (3) (b)
Gender (female)	0.00988	-0.0359	-0.0324	-0.0985*
	(0.035)	(0.052)	(0.045)	(0.057)
Age in adolescence (in months)	0.125***	0.140***	0.118***	0.104***
	(0.010)	(0.030)	(0.017)	(0.037)
Height-for-age z-score	0.0352***	0.111**	0.0409**	0.0597
	(0.013)	(0.049)	(0.019)	(0.059)
Constant	-1.519***		-1.398***	
	(0.13)		(0.22)	
Observations	279	279	172	172
Number of hh		131		82
R-squared	0.38	0.12	0.34	0.16

Notes: 1) Robust standard errors in parentheses. * Significant at 10%; ** Significant at 5%; *** Significant at 1%. 2) Standard errors for all the estimates are robust to clustered (village) sample design. 3) FE-IV are estimated using a linear probability model.

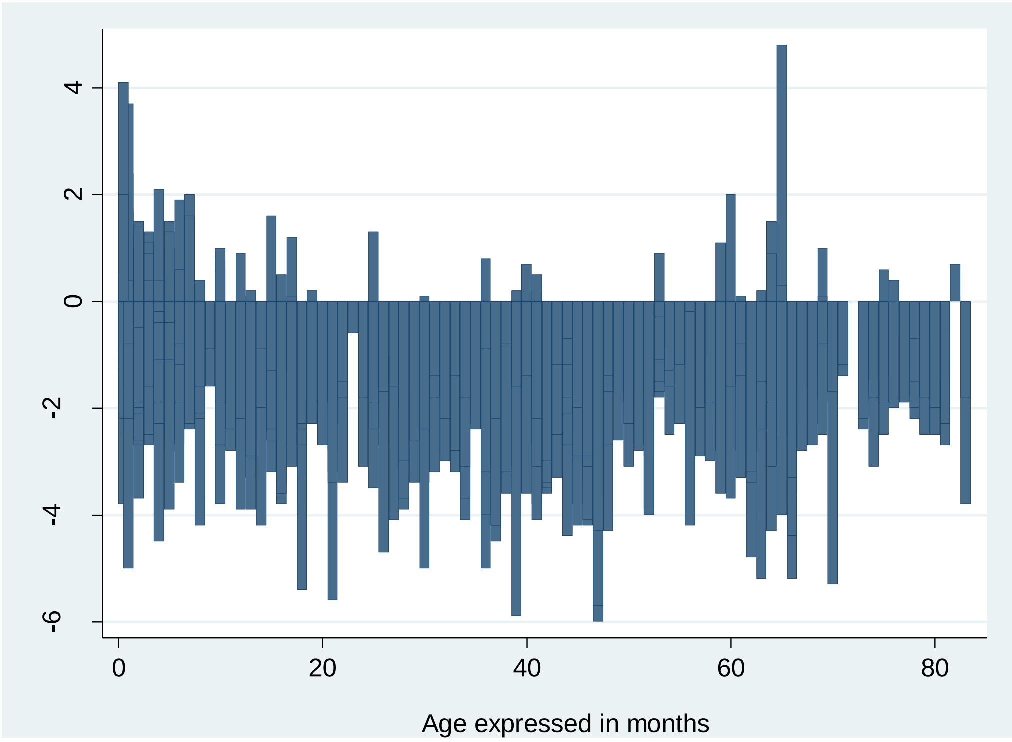
9. List of Figures

Figure 1: Height-for-age z-scores in Kagera, for pre-schoolers (1991-1994).



Source: Author's elaboration from KHDS data set

Figure 2: Height-for-age z-scores for pre-schoolers in Kagera, by age expressed in months (1991-1994).



Source: Author's elaboration from KHDS data set

Appendix: Health status for children living in Tanzania

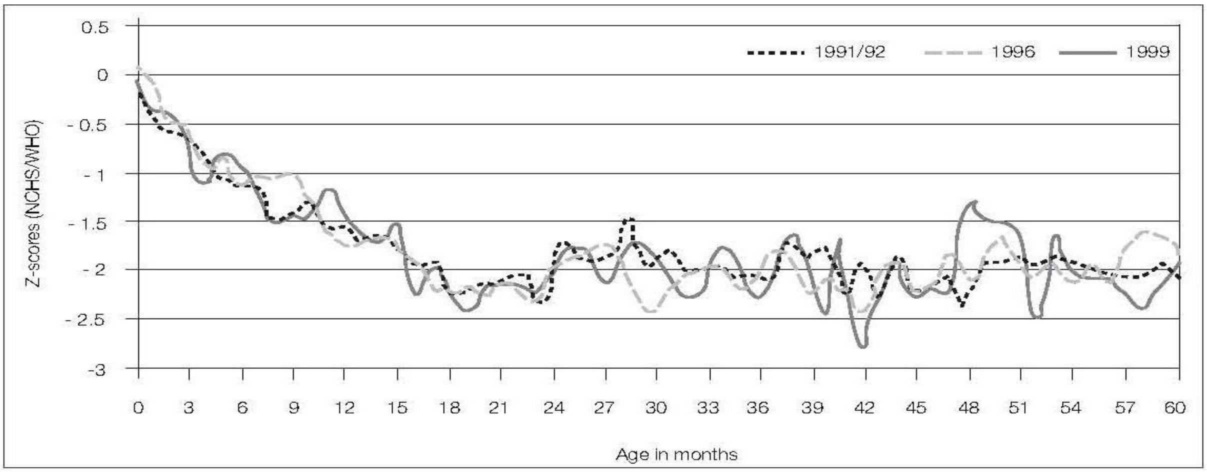
Deficiency of nutrition security is reflected in malnutrition affecting many Tanzanian children. A study of REPOA (2009) on nutrition status among children in Tanzania in the Nineties shows a pattern similar to that in the present study of Kagera (see Table 3, Figure 1 and Figure 2).

Table 16: Health status for children under-five years old living in Tanzania, represented by measure of stunting.

Variable	1991-92	1996	1999	2004-5
Height-for-age z-score<-2 SD	43%	43%	44%	38%

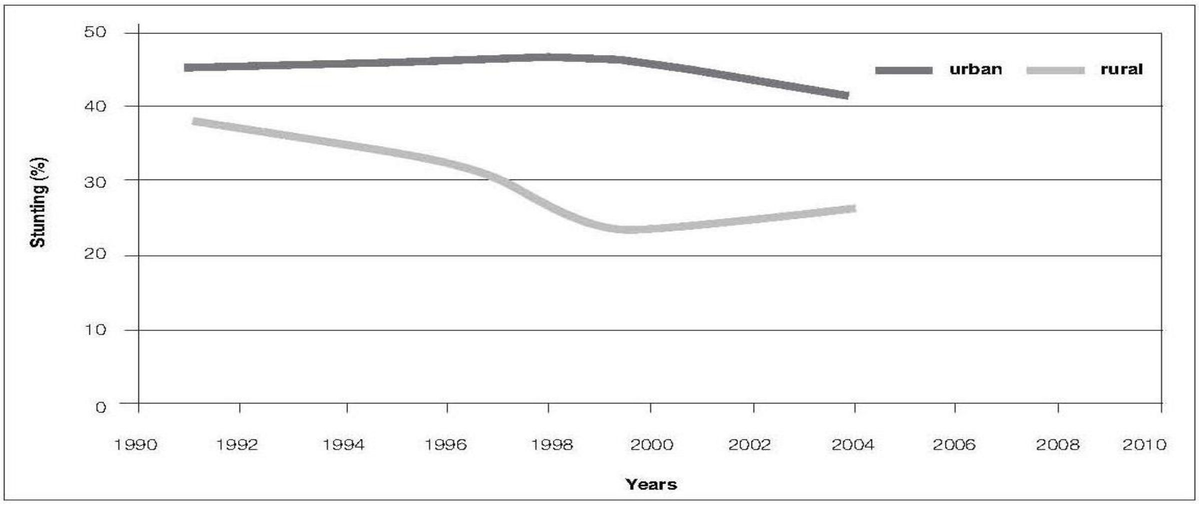
Source: REPOA (2009), calculated using TDHS 1991/92, TDHS 1996 and TRCHS 1999, TDHS 2004/2005.

Figure 3: Health status for children under-five years old living in Tanzania, represented by measure of stunting, by age in months, 1991-1999.



Source: REPOA (2009) and Lindeboom and Kilama (2005), calculations based on data from TDHS 1991/92, 1996 and 1999.

Figure 4: Prevalence of stunted children in urban and rural Tanzania, 1991-2004.



Source: REPOA (2009) and Lindeboom and Kilama (2005), using data from TDHS 1991/92-2004/05