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

Lucia Luzi, PhD
UNICEF IRC
lucia.luzi@unive.it, *lluzi@unicef.org*

2nd IRDES Workshop on Applied Health Economics and Policy Evaluation

June 23-24th 2011, Paris

ahepe@irdes.fr – www.irdes.fr

Outline

- Research question
- Motivations and contributions
- The econometric problem
- Results
- Final considerations and policy implications

- **Key words**: Primary education, child health and nutrition, weather shocks, family fixed effects, instrumental variables, Tanzania.

- **JEL classification**: I0

Research question

- *What are the effects of early childhood malnutrition on subsequent educational attainment in rural Tanzania?*

Motivations

- According to medical research the first 3 years of life are crucial for individual development.
- Exogenous shocks may cause permanent damage to children.
- Chronic malnutrition receives less policy attention than severe malnutrition, though prevalent in poor countries.

Contributions

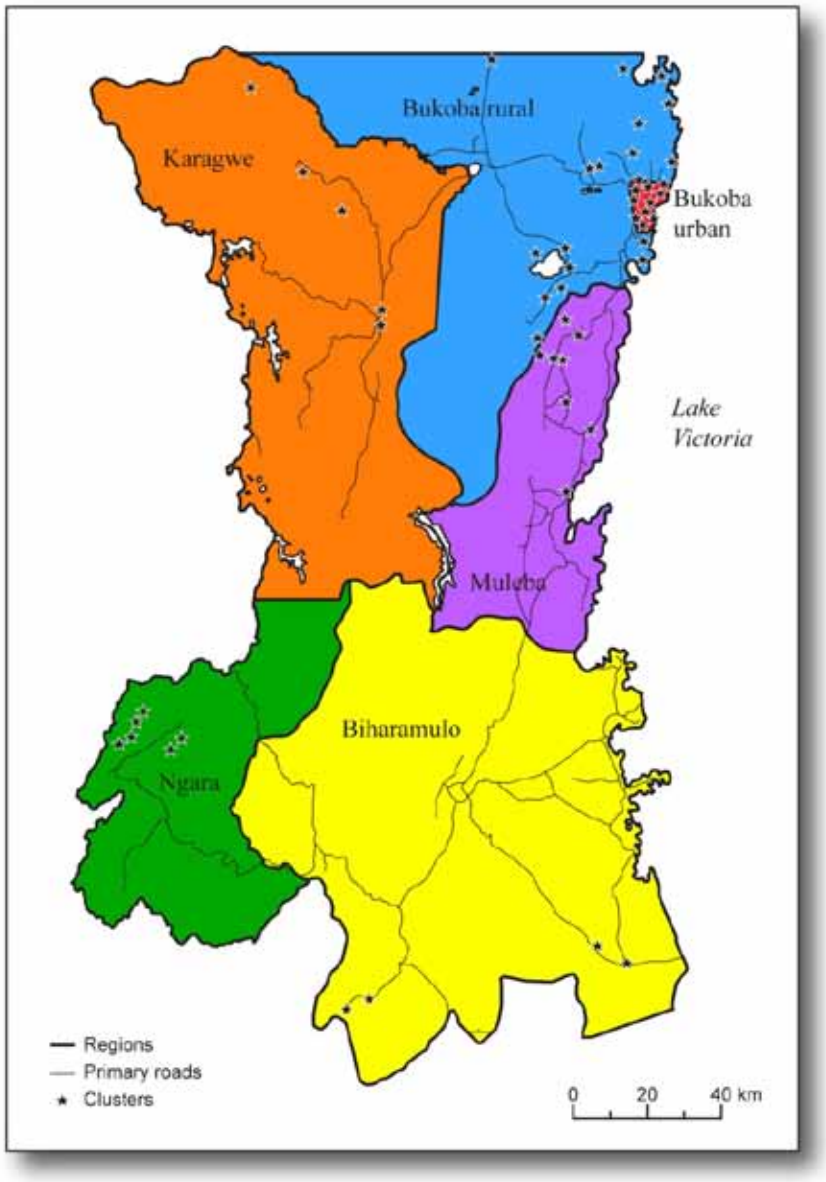
- This study:
  - extends the literature on the determinants of human capital formation in developing countries;
  - measures the impact of shocks at the individual level;
  - reveals aspects similar to other sub-Saharan African countries.


UGANDA
Lake Victoria
KENYA
Kagera
RWANDA
Mara
Arusha
BURUNDI
Mwanza
Kilimanjaro
Shinyanga
Kigoma
Manyara
Pemba North
Tanga
Tabora
Zanzibar North
Pemba South
Lake Tanganyika
Singida
Dodoma
Zanzibar Urban/West
Zanzibar Center/South
TANZANIA
Rukwa
Morogoro
Dar es Salaam
DEM.REP. OF THE CONGO
Iringa
Pwani
Mbeya
INDIAN OCEAN
Lindi
ZAMBIA
Ruvuma
Mtwara
MALAWI
Lake Nyasa
MOZAMBIQUE

Kagera Health & Development Survey

- **Dataset:**
  - Kagera Health and Development Survey (LSMS)
  - Conducted by the World Bank, Muhimbili University College of Health Sciences and University of Dar es Salaam.
- **Periods:** 4 times 1991-1994 (KHDS I)

  + 1 time 2004 (KHDS II)
- **Population:** 915 hh drawn from 51 communities of 16 hh each in the 6 administrative districts of Kagera.
- **Advantages:**
  - it is one of the few surveys that has data over such a long period;
  - it has a low attrition rate of 9,6%;
  - it particularly appropriate for the analysis.



Kagera Region, Tanzania. KHDS clusters' location.

The population was 1.3 mln in 1988, and about 2 mln in 2004. It is overwhelmingly rural and primarily engaged in producing bananas and coffee in the northern districts and rain-fed annual crops (maize, sorghum, cotton) in the southern districts.

Literature

- This study follows previous elaborations made by:
  - Glewwe, Jacoby and King (2001)
  - Alderman, Behrman, Lavy and Menon (2001)
  - Alderman, Hoddinott and Kinsey (2006)
  - Glewwe and Miguel (2008)

The econometric model

- Two time periods model:
  - t=1 the individual is a newborn or a preschooler (KHDS I)
  - t=2 the individual is an adolescent or a young adult (KHDS II)

- In each period parents make decisions on child's human capital investments, but those in t=1 are the most important with long-term effects.

The econometric model

The structural form:

$$S_{i2} = a_H f(H_{i1}) + a_{C2} g(C_{i2}) + \varepsilon_{i2} \qquad (1)$$

- i is the identification for the child
- S_{i2} is the educational outcome of the child i at t=2
- $f(H_{i1})$ is a function of health status of the child i at t=1
- C_{i2} is a vector of individual, hh and community characteristics that influence academic performance
- ε_{i2} is the individual specific disturbance term that affects the educational outcome of interest

The econometric model

The reduced form:

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

- H_{i1} is the health status of the child i at t=1
- C_{i1} is a vector of individual, hh and community characteristics that influence investment in health
- ε_{i1} is the individual specific disturbance term that affects the health status

Endogeneity problem

- OLS method can produce biased estimates since it:
  - requires the availability of complete data on all the right hand var. in eq.(1), while some factors are unobserved;
  - assumes that H_{i1} is exogenous (pre-determinate), while it is endogenous and probably correlated with ε_{i2} : $E(H_{i1}\ \varepsilon_{i2}) \neq 0$. This can be caused by possible correlations of individual or hh effects, unobservable by the data analyst.

➡ In performing such analysis an **endogeneity problem** exists.

Tackling the endogeneity problem

1. The **within-sibling** approach (FFE) purges any hh and environment inputs (both observed and unobserved) that are constant across siblings.
2. The **instrumental variable** approach (IV) purges any unobserved correlations of individual effects.

H_{i1} is first estimated using IV_{i1} and then S_{i2} is estimated using $\hat{H}_{i1}$ from the first stage.

Weather shock as IV

- IV = R_{i1} : weather shock at location and time of birth for each child. The shock takes place after parents have made decisions for that time period.

- As IV, R_{i1} is:
  - of adequate magnitude and persistence to affect H_{i1}
  - adequately variable across siblings in the same hh
  - adequately transitory not to affect H_{k1}
  - not correlated with S_{i2}

- R_{i1} satisfy the two conditions of:
  1. *Instrument relevance:* $\text{cov}(R_{i1}, H_{i1}) \neq 0$
  2. *Instrument exogeneity:* $\text{cov}(R_{i1}, \varepsilon_{i2}) = 0$

Variables and measures

- The suitable sample for the analysis is constituted by children with available information on:
  - H_{i1} measured by height-for-age
    - A low height-for-age z-score defines "stunting", which indicates chronic malnutrition
  - S_{i2} measured by completion of the entire cycle of primary education
  - R_{i1} measured by rainfall at location and time of birth

Table 1: Descriptive statistics on children in KHDS '91-'94

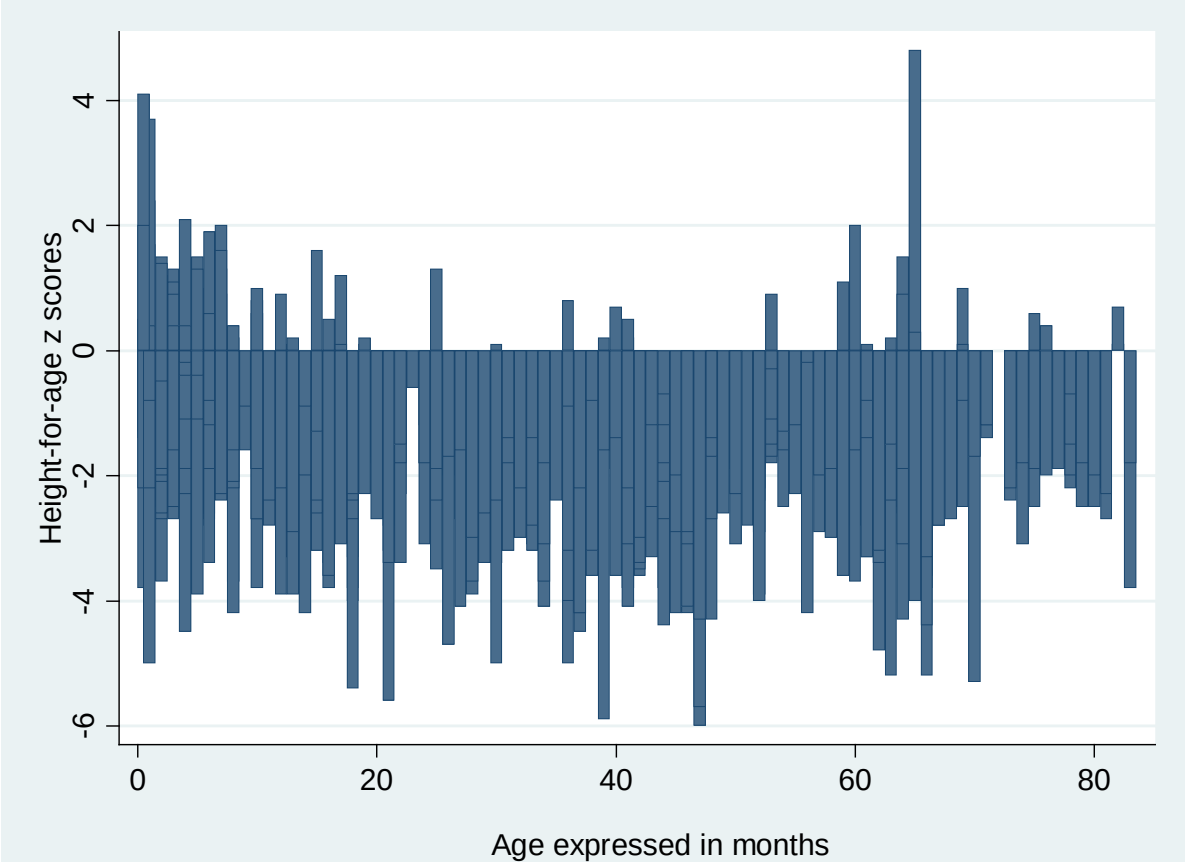
Variable	Obs	Mean	Std. Dev.
Height-for-age z-score	622	-1.65	1.50
Stunted	622	0.70	0.46
Age (in months)	622	32.58	24.87
Gender (female)	622	0.46	0.50

Table 2: Heath status of children in KHDS '91-'94

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%

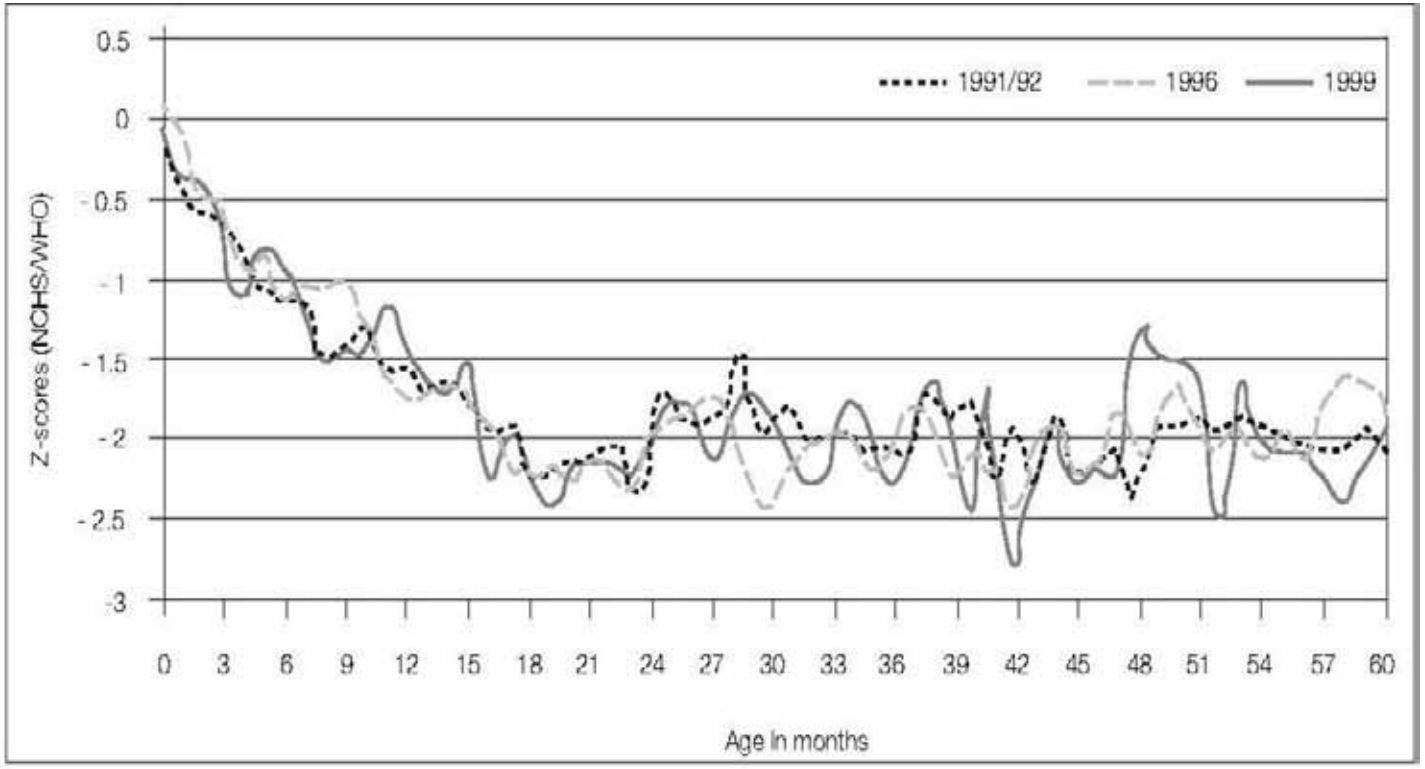
Source: Author's elaboration from KHDS dataset

Figure 1: Height-for-age z-scores for pre-schoolers in KHDS '91-'04, by age expressed in months



Source: Author's elaboration from KHDS dataset

Figure 2: Health status (stunting) for children under-5 years old in Tanzania, in months, '91-'99



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

Table 3: 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:

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors for all the estimates are robust to clustered (village) sample design. FFE-IV are estimated using a linear probability model. 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 3: 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:

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors for all the estimates are robust to clustered (village) sample design. FFE-IV are estimated using a linear probability model. 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.

Plausible reasons for statistically significant height-for-age in the selected sub-sample

- Karagwe (district n.1) and Ngara (district n.5):
  - have the worst health performance on average;
  - are the driest areas, located far from Lake Victoria;
  - were the primary asylum for the refugees from Burundi and Rwanda genocides to escape ethnic violence during the early '90, with consequent damages.

Table 4: First-stage within siblings regression (sub-sample)

Estimation Approach	Instrumental Variables	Family Fixed Effects - Instrumental Variables (3)
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%

2. Standard errors for all the estimates are robust to clustered (village) sample design.

3. "Family Fixed Effects - Instrumental Variables" are estimated using a linear probability model

Table 4: First-stage within siblings regression (sub-sample)

Estimation Approach	Instrumental Variables	Family Fixed Effects - Instrumental Variables (3)
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%

2. Standard errors for all the estimates are robust to clustered (village) sample design.

3. "Family Fixed Effects - Instrumental Variables" are estimated using a linear probability model

Table 5: Estimates of the education achievement equation for siblings (sub-sample)

Estimation Approach	OLS (1)	FFE-IV(2)	FFE-IV(3)	OLS (4)	FFE-IV(5)
Gender (female)	0.0171	-0.0212	-0.154*	-0.0360	-0.152*
	(0.031)	(0.047)	(0.080)	(0.044)	(0.079)
Age in adolescence	0.117***	0.129***	0.145***	0.109***	0.143***
(in months)	(0.0072)	(0.012)	(0.030)	(0.013)	(0.030)
Height-for-age	0.0451***	0.125**	0.192*	0.0271	0.189*
z-score	(0.011)	(0.051)	(0.10)	(0.020)	(0.10)
Constant	-1.403***			-0.946	
	(0.091)			(0.80)	
Controls	No	No	No	Yes	Yes
Observations	450	450	254	254	254
Number of hh		173	102		102
R-squared	0.38	0.36	0.16	0.34	0.18

Notes:

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors for all the estimates are robust to clustered (village) sample design. FFE-IV are estimated using a linear probability model. Controls at individual, hh and community level (OLS) – Dummy for vaccine only (FFE-IV). (3) represents the analysis on sample (4) and (5) without controls.

Table 5: Estimates of the education achievement equation for siblings (sub-sample)

Estimation Approach	OLS (1)	FFE-IV(2)	FFE-IV(3)	OLS (4)	FFE-IV(5)
Gender (female)	0.0171	-0.0212	-0.154*	-0.0360	-0.152*
	(0.031)	(0.047)	(0.080)	(0.044)	(0.079)
Age in adolescence	0.117***	0.129***	0.145***	0.109***	0.143***
(in months)	(0.0072)	(0.012)	(0.030)	(0.013)	(0.030)
Height-for-age	0.0451***	0.125**	0.192*	0.0271	0.189*
z-score	(0.011)	(0.051)	(0.10)	(0.020)	(0.10)
Constant	-1.403***			-0.946	
	(0.091)			(0.80)	
Controls	No	No	No	Yes	Yes
Observations	450	450	254	254	254
Number of hh		173	102		102
R-squared	0.38	0.36	0.16	0.34	0.18

Notes:

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Standard errors for all the estimates are robust to clustered (village) sample design. FFE-IV are estimated using a linear probability model. Controls at individual, hh and community level (OLS) – Dummy for vaccine only (FFE-IV). (3) represents the analysis on sample (4) and (5) without controls.

Results and Final Considerations

- Applying the FFE – IV approach, a Tanzanian child in good health status during infancy has almost an additional 28% probability (=0.189*average H_{i1}) of completing primary education.
- Policy implications: Investing in education and health is critical for the future; hence, it should be a priority for governments and policy makers.
- Improvements in health status and primary education are not competing goals, but mutually reinforcing.
- Long-run effects of early-life conditions on schooling should be factored into cost-benefit analyses of government programs.

Thank you for your kind attention!