Analysis of Microdata: Tanzania
Policy Evaluation and Applied Statistics

ROMAN KÜPPER
MAS Development and Cooperation NADEL ETHZ 20-22
December 22, 2020

1 Introduction

While stunting and mortality for children under five have declined in most countries over the last decades, they are still important issues for countries in sub-Saharan Africa [8]. The reduction of child mortality is also indirectly linked to reduced fertility due to the higher chances of a child's survival [9]. Furthermore stunting and under-five mortality are and have been important topics in development for the last decades. The reduction of child mortality is the stated aim of Goal 4 of the Millennium Development Goals (MDGs) and the health of children is mentioned in several of the Sustainable Development Goals (SDGs). To successfully lower the stunting and child mortality rates, it is vastly important to understand the underlying causes. While economic growth is a common approach in development to reach better child health, recent studies have shown that there are only small to null associations between income growth and children's health factors like undernutrition [11]. The following article will analyse data from the Demographic and Health Survey (DHS) for Tanzania in order to identify country-specific factors that determine if a child is at the risk of being stunted or prone to die below the age of five. The study will focus mainly on the effects of per-household income and examines if income is a necessary and sufficient condition for improvements in child health.

Using the DHS dataset for Tanzania from 2005 to 2015, multiple regression models are built to find out how per household income is related to child stunting and under-five mortality. The findings suggest that income has a significant effect on stunting but no significant effect on the under-five mortality rate of children.

Tanzania

The situation of Tanzanian children regarding stunting and under-five mortality has improved over the last three decades. The prevalence for stunting could be decreased from 50% in the 1990s to 34% in the year 2015 [7]. The same applies to the under-five mortality rate. The sex-specific under-five mortality rate dropped from 171 to 54 per 1000 births for boys and from 159 to 47 per 1000 births for girls from 1990 to 2019 [10]. But despite improvements in recent years, stunting and child mortality are still ongoing issues in Tanzania.

Other literature

Previous research has shown several different factors that influence stunting and under-five mortality. These include the age and sex of the child and various factors concerning the mother's health and education. But also socio-economic factors like household wealth are mentioned. The factors mentioned in the literature differ for stunting and under-five mortality. Unfinished listings for both, stunting and under-five mortality, can be found in the according tables in Appendix A and Appendix B. The tables in the appendix also show how the factors are matched with the variables in our dataset. Factors often mentioned as related to stunting are child age, sex of the child, birth weight, birth size, mothers health, mothers education, mothers age, mothers BMI, breastfeeding, wealth of household, social inequality, source of drinking water, sanitation and place of residence [2][7].

Some of the factors related to stunting are also mentioned as factors for under-five mortality, which is not surprising since the most severe forms of stunting often lead to death. Literature lists sex of the child, birth order, malnutrition, vaccinations, access to postnatal care, mothers age, education, breastfeeding, place of delivery, household wealth, place of residence, source of drinking water and sanitation as some of the most important factors for under-five mortality according to [4][12] and [5].

2 Methods

Data sources

In order to analyse the relationship between income and under-five mortality and stunting rate, a dataset based on the Demographic and Health Surveys (DHS) [1] by ICF International is used. The dataset contains panel data

of Tanzanian households for the years 2005, 2010 and 2015. The dataset consists of 24,198 observations for 15,273 households.

Research Strategy

This research aims to provide a better understanding of the effectiveness of income on under-five-mortality and stunting. For this reason multiple regression models are built to estimate the influence of different independent variables on the two dependent variables (stunting and under-five-mortality). In a first step, only the influence of income (X_i) on stunting and under-five-mortality(Y_i) is examined. The formula for the traditional ordinary least squares model (1) is shown below:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \tag{1}$$

Next, the models are adjusted for survey-year time-fixed effects (2) with the formula:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_T BT_t + u_{it} \tag{2}$$

The models are then further improved by adding other potentially meaningful variables (3) to further reduce the error term (u_{it}). The formula for the multiple regression with regard to time-fixed effects (δ_T) and binary variables (B_t) in the "T-1 binary regressor"[1] formulation is the following:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_t + \cdots + \delta_T BT_t + u_{it} \tag{3}$$

Starting from the simple ordinary least square (OLS) regression, the model will be further enhanced by controlling for additional variables and time fixed effects.

Data sources and procedures

To provide a better understanding of the dataset, descriptive statistics of the most interesting variables are shown in this section. A tabular overview can be found in table 1. Some information is further presented in graphs to allow a better visual interpretation. The dataset contains information from surveys in 2005, 2010 and 2015. The whole number of children included in the dataset is 24,198. The number of children has declined slightly from the first survey in 2005 with 7,727 observations in 2005 to 7,303 in 2010 but has then inclined to 9,168 in 2015.

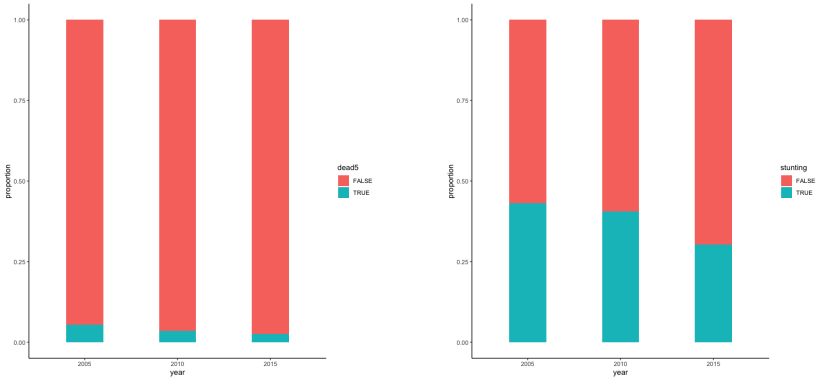
[1] More information on this can be found in the slides of the course "Policy Evaluation and Applied Statistics" on Panel Regressions.

	2005 (N=7727)	2010 (N=7303)	2015 (N=9168)	Overall (N=24198)
Stunting				
no	3775 (48.9%)	3772 (51.7%)	5736 (62.6%)	13283 (54.9%)
yes	2854 (36.9%)	2581 (35.3%)	2501 (27.3%)	7936 (32.8%)
missing	1098 (14.2%)	950 (13.0%)	931 (10.2%)	2979 (12.3%)
under-five years mortality				
no	7303 (94.5%)	7050 (96.5%)	8928 (97.4%)	23281 (96.2%)
yes	424 (5.5%)	253 (3.5%)	240 (2.6%)	917 (3.8%)
Under-1 year mortality				
no	1664 (21.5%)	1541 (21.1%)	1887 (20.6%)	5092 (21.0%)
yes	282 (3.6%)	167 (2.3%)	144 (1.6%)	593 (2.5%)
missing	5781 (74.8%)	5595 (76.6%)	7137 (77.8%)	18513 (76.5%)
Under-1 month mortality				
no	69 (0.9%)	68 (0.9%)	82 (0.9%)	219 (0.9%)
missing	7658 (99.1%)	7235 (99.1%)	9086 (99.1%)	23979 (99.1%)
Household income				
Mean (SD)	66.6 (56.5)	78.0 (60.1)	82.2 (68.1)	76.0 (62.5)
Median [Min, Max]	48.2 [6.43, 485]	60.3 [13.2, 438]	62.7 [11.2, 426]	60.3 [6.43, 485]
Place of residence				
rural	6383 (82.6%)	5918 (81.0%)	7034 (76.7%)	19335 (79.9%)
urban	1344 (17.4%)	1385 (19.0%)	2134 (23.3%)	4863 (20.1%)
Child age (months)				
Mean (SD)	28.2 (17.6)	29.1 (17.6)	29.0 (17.5)	28.8 (17.6)
Median [Min, Max]	27.0 [0, 60.0]	29.0 [0, 60.0]	28.0 [0, 60.0]	28.0 [0, 60.0]
Child sex				
male	3846 (49.8%)	3626 (49.7%)	4590 (50.1%)	12062 (49.8%)
female	3881 (50.2%)	3677 (50.3%)	4578 (49.9%)	12136 (50.2%)
Water: improved				
no	4048 (52.4%)	3822 (52.3%)	3999 (43.6%)	11869 (49.0%)
yes	3679 (47.6%)	3481 (47.7%)	5169 (56.4%)	12329 (51.0%)
Sanitation: improved				
no	7328 (94.8%)	5841 (80.0%)	3068 (33.5%)	16237 (67.1%)
yes	399 (5.2%)	1462 (20.0%)	6100 (66.5%)	7961 (32.9%)
Mothers age (years)				
Mean (SD)	29.1 (6.80)	29.7 (6.98)	29.5 (7.13)	29.4 (6.99)
Median [Min, Max]	28.0 [15.0, 49.0]	29.0 [15.0, 49.0]	29.0 [15.0, 49.0]	29.0 [15.0, 49.0]
Mothers age (years) at first birth				
Mean (SD)	19.0 (3.10)	19.2 (3.12)	19.4 (3.34)	19.2 (3.20)
Median [Min, Max]	19.0 [10.0, 35.0]	19.0 [10.0, 41.0]	19.0 [10.0, 46.0]	19.0 [10.0, 46.0]
Mothers total children				
Mean (SD)	4.18 (2.55)	4.26 (2.52)	4.06 (2.57)	4.16 (2.55)
Median [Min, Max]	4.00 [1.00, 14.0]	4.00 [1.00, 15.0]	3.00 [1.00, 17.0]	4.00 [1.00, 17.0]
Mothers breastfeeding status				
no	3017 (39.0%)	2926 (40.1%)	3789 (41.3%)	9732 (40.2%)
yes	4710 (61.0%)	4377 (59.9%)	5379 (58.7%)	14466 (59.8%)
Mother received prenatal care				
no	4230 (54.7%)	3344 (45.8%)	8705 (94.9%)	16279 (67.3%)
yes	3497 (45.3%)	3959 (54.2%)	463 (5.1%)	7919 (32.7%)
Mother received medical assistance				
no	4584 (59.3%)	3878 (53.1%)	8323 (90.8%)	16785 (69.4%)
yes	3143 (40.7%)	3425 (46.9%)	845 (9.2%)	7413 (30.6%)
Mother education				
none	3390 (43.9%)	2986 (40.9%)	3208 (35.0%)	9584 (39.6%)
primary	4235 (54.8%)	4225 (57.9%)	5052 (55.1%)	13512 (55.8%)
secondary	102 (1.3%)	92 (1.3%)	908 (9.9%)	1102 (4.6%)

Table 1: General descriptive statistics of baseline characteristics of the study population

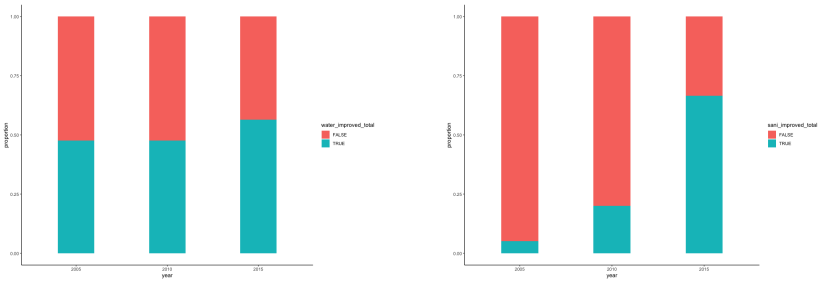
Outcomes

The share of children that are stunted as well as the share of children dying under five has declined through the three-time periods. A visualization of this trend can be seen in figure 1 which shows the proportions of both variables over time. The figure also shows that while child mortality has been vastly reduced, stunting is still a widely spread issue in Tanzania.



(a) Under-five mortality (b) Stunting

Figure 1: Proportions of children that are stunted or have died under the age of fiver over time



(a) Access to improved water (b) Access to improved sanitation

Figure 2: Proportions of children with access to improved water and sanitation over time

Figure 3 shows the per household income from 2005 to 2015 grouped by location (urban or rural). The graph shows large differences between urban and rural over all time periods. This income gap between cities and the countryside is in line with existing literature. It is also noticeable that the differences between households in both locations are quite large. The average household income has slightly increased from 2005 to 2010 for both

locations and was almost constant from 2010 to 2015.

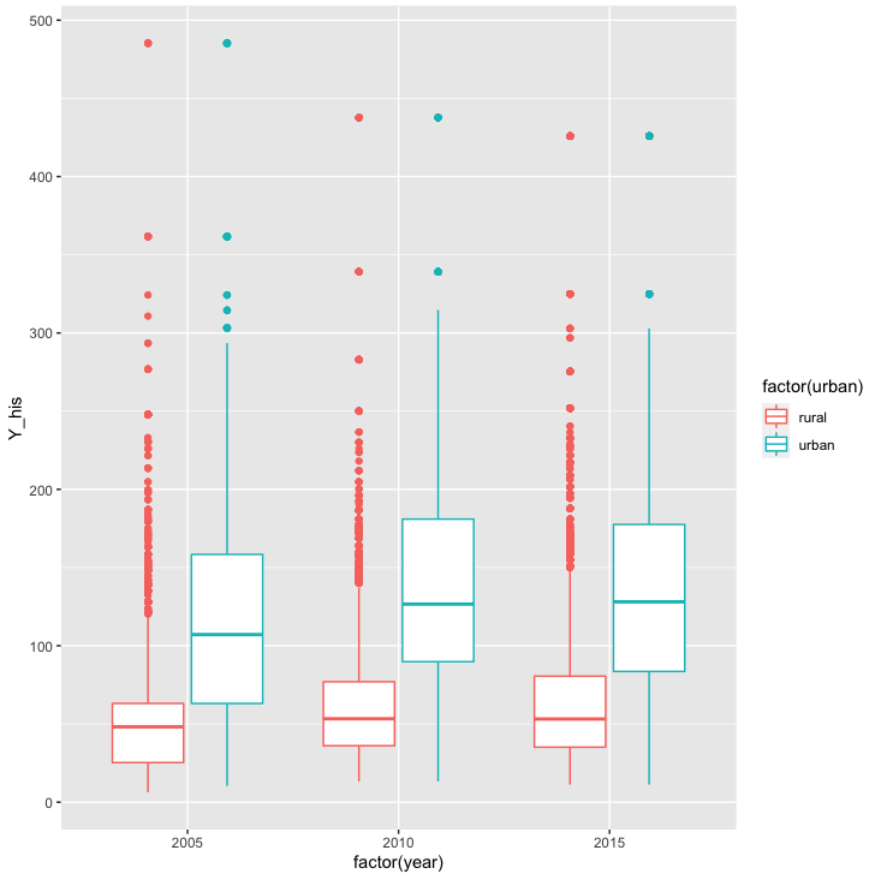
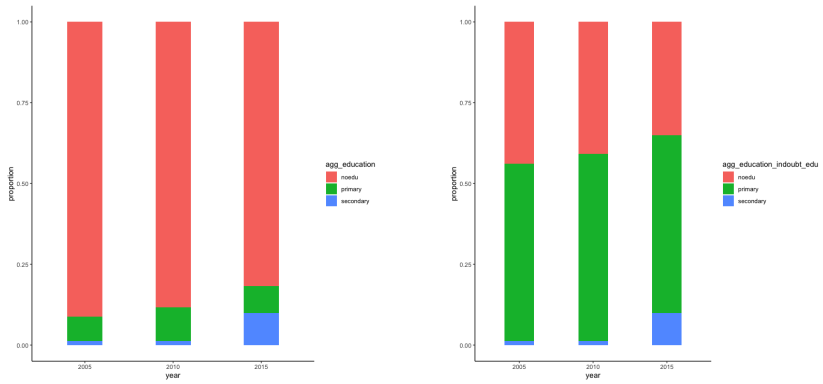


Figure 3: Y_his over time

Access to clean water sources and improved sanitation are also substantively important variables for the well-being of children. If they are lacking they can cause life-threatening diseases like diarrhea that are under suspicion to cause up to 50% of all child malnourishment [7]. The share of children with access to improved water and sanitation is shown in figure 2. The figures for improved water as well as improved sanitation show an upward trend with the share of improved sanitation (see figure 2(b)) having climbed more rapidly than the share of access to improved water (see figure 2(a)). The trend towards better water and sanitation is in line with the Tanzanian government's vision for a more healthy society from 2007 [6].

Another often quoted factor for the well-being of children is their mother's education. The dataset includes information on three different variables for education. Namely primary, secondary and no education. The interpretation of the data is somewhat difficult. Some of the observations specify that a mother has both primary and no education at the same time. When this data is interpreted as "the mother has started but not finished her primary education", the corresponding barplot is shown in figure (4)(a).



(a) Lower education if in doubt (b) Higher education if in doubt

Figure 4: Mothers education levels over time

But when looking at other studies on the distribution of primary, secondary on no-education in Tanzania, the share of uneducated women seems to be too high. If the ambiguous values are interpreted as "the education is not finished yet, but will be"[2], our data is more in line with other studies [3]. The graphic representation of this can be found in figure (4)(b).

[2]When an observation has "mo_noedu == 1" and "mo_primary== 1" the variables are aggregated to "education = primary".

Table 2: Regressions Using Demographic and Health Surveys

	Dependent variable:					
	stunting			dead5		
	(I)	(II)	(III)	(IV)	(V)	(VI)
	OLS	Fixed effects	Full	OLS	OLS fixed effects	Full
log_y	−0.10287***	−0.09724***	−0.06590***	−0.00832***	−0.00632***	−0.00354
	(0.00431)	(0.00433)	(0.00568)	(0.00162)	(0.00161)	(0.00222)
as.factor(urban)urban			−0.04095***			0.00569*
			(0.00891)			(0.00341)
c_age			0.00288***			
			(0.00020)			
c_sex			−0.05025***			
			(0.00645)			
c_first						0.02076***
						(0.00370)
mo_assistance						−0.00630**
						(0.00314)
mo_age_birth			−0.00137***			0.00180***
			(0.00048)			(0.00023)
as.factor(mo_breastfeeeding)yes			−0.04524***			−0.02895***
			(0.00757)			(0.00267)
mo_primary			−0.03265***			−0.00395
			(0.00707)			(0.00272)
mo_secondary			−0.06947***			−0.01304***
			(0.01342)			(0.00442)
water_improved_total			−0.02413***			−0.00982***
			(0.00710)			(0.00273)
sani_improved_total			−0.02135**			−0.00636**
			(0.00908)			(0.00317)
Constant	0.79150***	0.81154***	0.78864***	0.07170***	0.07966***	0.04294***
	(0.01817)	(0.01828)	(0.02868)	(0.00689)	(0.00710)	(0.01069)
Observations	21,219	21,219	21,219	24,198	24,198	24,198
R^2	0.02397	0.03486	0.05857	0.00102	0.00462	0.01457
Adjusted R^2	0.02392	0.03472	0.05803	0.00098	0.00449	0.01408

Note: *p<0.1; **p<0.05; ***p<0.01

3 Results

The variables included in the models are chosen based on the available dataset, the existing literature and the author's limited knowledge on the matter. To further get insights on variables with a significant impact on both outcome variables, a series of correlation matrices was created. The final matrices can be found in Appendix 5 and Appendix 6. The circles in each matrix show how each set of variables on X and Y-axis are correlated. The color of a circle determines how strong the correlation is and if it is positive or negatively correlated. Colour intensity and the size of the circles are proportional to the correlation coefficients. The analysis of the two correlation matrices already gives a first impression of the effects of the different independent variables on our outcome variables. While we won't go into detail with the interpretation of the matrices, it is interesting to see how correlated some of the variables are. The findings from the visual analysis will go into the creation of our regression models presented in this section.

In the first step of the statistical analysis, an ordinary least squares regression model is chosen to estimate the effects of income-changes on the binary outcome variables stunting and under-five mortality. To allow a better interpretation, the natural logarithm of a household's income (log_y) is chosen. The overall strategy is to judge the influence of a per household income change on a child's probability to be stunted or to die under the age of five while controlling for time-fixed effects and other influential variables. The control variables differ between stunting and under-five mortality. Not the whole dataset contains information on stunting. Observations that have missing values on stunting are omitted for the regression model that is built for stunting. This leads to a reduction of observations for the models (I-III) compared to the models (IV-VI) from 24,198 to 21,219.

The findings are listed in table 2 (I-III) for stunting and (IV-VI) for under-five mortality. In column (I) for stunting and column (IV) for under-five mortality, we can see the results of the most simple OLS regression. As our outcome variables are binary (stunting / not stunting and dead under-five and alive under-five) and our independent variable is the natural logarithm of the household income, the effect is estimated based on percentage changes in income. For stunting, this means that an additional percent of income would lead to a by 0.1 percentage point decreased probability of a child being stunted. For under-five mortality, this effect is a lot smaller and the additional percent of income would only lead to a decrease of 0.008 percentage points. Both effects are significant to a

significance level of 5%.

In section 2 it was shown that stunting and under-five mortality rates have been declined over time. Since it cannot be certain that all explaining variables have been included in our model, it is important to control for time-fixed effects. They are included to control for the fact that we can expect the number of stunting and dying children to be declining based on factors we cannot control for that are changing over time. If those time-fixed effects are included, the effect of additional income is shrinking slightly. This could be due to the long-term results of the vision for a more healthy society in Tanzania or better access to health-related articles, a better understanding of the matter, or other effects that aren't included in the model or where no data is available. But even with time-fixed effects included, both effects, while being small, are still significant.

The final models are shown in (III) and (VI) of table 2. For stunting, the additional variables on the place of residence, child age, child sex, mother's age at birth, mother is breastfeeding, mothers education, water and sanitation are included. With these variables included, the effect that is explained by additional income further decreases to 0.007 percentage points. All of the variables mentioned have a significant effect, but the effect of sanitation is only significant to a lower significance level of 10%. In total 0.003 percentage points of the original effect could be explained by time fixed effects and other variables. But even with all those variables included, household income still has the second-highest impact on the probability of a child being stunted. Other important factors are if the mother has a secondary education, the child's sex, if the mother is breastfeeding and place of residence (urban).

For under-five mortality, the additional variables on the place of residence, first-born child, mother got assistance, mother got care, mother's age at birth, mother is breastfeeding, mothers education, water and sanitation are included. With these additional variables included, the effect of income further declines and is not significant anymore. The three variables with the largest effect on under-five mortality are if the mother is breastfeeding, if the child is the first-born and if the mother got secondary education. Water and sanitation also have measurable effects on under-five mortality. Since both of those correlate with income, it could be that the higher income already explains a large share of the income effects. Also, the comparatively high effect of a child being the first-born could be interpreted as lacking knowledge or experience of a mother that could lead to a higher

chance of the child to die. But to answer our initial question, it seems as if income is not a sufficient condition for a smaller risk of a child to die under the age of five.

Since our model estimates the effect on a binary dependent variable, both R^2 and Adjusted-R^2 do not provide meaningful information on the goodness of fit of the model and will therefore not be interpreted.

4 Discussion

In the previous chapters, it was shown how household income affects child health in Tanzania. More exactly it was examined how stunting and under-five mortality rates are linked with income and what impact other variables like the mother's health and the state of sanitation and water access have an impact. The analysis provides mixed effects on our outcome variables. For stunting, it could be shown that income changes could have a significant impact even when taking other variables and time-fixed effects into account. This implies that if household incomes in Tanzania grow further, stunting rates of children could be further reduced.

But a growing per-household income alone is not sufficient in reducing under-five mortality. The findings provided in the previous section suggest that it is more effective to invest in better prenatal care or provide information on the benefits of breastfeeding for children's health [3].

There are some limitations to the presented model, since some of the variables suggested to have an impact by previous research could not be included. This could lead us to tend to evaluate the impact of income as too large. Also, it has to be kept in mind that some of the variables used in the models (III) and (VI) are intercorrelated as shown in the correlation matrices in Appendix C and D. An enhanced model should therefore explicitly test for multicollinearity by including a variance inflation factor (VIF). Furthermore, some variables of interest from the dataset weren't added due to missing data. These include under-one-month- and under-one-year mortality and mothers without education.

[3] The positive effects of breastfeeding are broadly mentioned in literature. If children are only fed over a mothers breast, the chances of drinking or eating contaminated water or food are small and children are less likely to get diseases and infections.

5 Acknowledgements

Thanks go out to the other statistical novices at NADEL with special thanks to Gabriel for sharing his thoughts, David et al. for organizing our anonymous statisticians' meetings and Viola and Joelle for their valuable inputs. I would further like to thank Ken for providing us with this interesting dataset and a real-world example.

References

1. The DHS Program - Data. `https://dhsprogram.com/data`, Dec 2020. [Online; accessed 11. Dec. 2020].

2. Blessing J. Akombi, Kingsley E. Agho, John J. Hall, Nidhi Wali, Andre M. N. Renzaho, and Dafna Merom. Stunting, Wasting and Underweight in Sub-Saharan Africa: A Systematic Review. *Int. J. Environ. Res. Public Health*, 14(8), Aug 2017.

3. Samer Al-Samarrai and Barry Reilly. Education, Employment and Earnings of Secondary School and University Leavers in Tanzania: Evidence from a Tracer Study. *Journal of Development Studies*, 44(2):258–288, Feb 2008.

4. Remare Ettarh and James Kimani. Determinants of under-five mortality in rural and urban Kenya, Mar 2012. [Online; accessed 19. Dec. 2020].

5. UNICEF (United Nations Children's Fund). Children and Water, Sanitation and Hygiene: The Evidence | Human Development Reports, Dec 2006. [Online; accessed 19. Dec. 2020].

6. Ministry of Health, Community Development, Gender, Elderly and Children - MoHCDGEC/Tanzania Mainland, Ministry Of Health-Moh/zanzibar, National Bureau Of Statistics-Nbs/tanzania, Office Of Chief Government Statistician-Ocgs/zanzibar, and Icf. Tanzania Demographic and Health Survey and Malaria Indicator Survey 2015-2016, Dec 2016. [Online; accessed 21. Dec. 2020].

7. Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC) [Tanzania Mainland], Ministry of Health (MoH) [Zanzibar], Tanzania Food and Nutrition Centre (TFNC), National Bureau of Statistics (NBS), Office of the Chief Government Statistician (OCGS) [Zanzibar] and UNICEF. Tanzania National Nutrition Survey using SMART Methodology 2018, 2018.

8. Alphonse Nshimyiryo, Bethany Hedt-Gauthier, Christine Mutaganzwa, Catherine M. Kirk, Kathryn Beck, Albert Ndayisaba, Joel Mubiligi, Fredrick Kateera, and Ziad El-Khatib. Risk factors for stunting among children under five years: a cross-sectional population-based study in Rwanda using the 2015 Demographic and Health Survey. *BMC Public Health*, 19(1):1–10, Dec 2019.

9. Jamal Uddin, Zakir Hossain, and Mohammad Ohid Ullah. Child mortality in a developing country: A statistical analysis. 2009.

10. United Nations Interagency Group for Child Mortality Estimation (UN IGME). Levels and Trends in Child Mortality: Report 2020, Estimates developed by the United Nations Inter-agency Group for Child Mortality Estimation, 2020.

11. Sebastian Vollmer, Kenneth Harttgen, Malavika A. Subramanyam, Jocelyn Finlay, Stephan Klasen, and S. V. Subramanian. Association between economic growth and early childhood undernutrition: evidence from 121 Demographic and Health Surveys from 36 low-income and middle-income countries. *Lancet Global Health*, 2(4):e225–e234, Apr 2014.

12. World Health Organization: Who. Children: improving survival and well-being. *World Health Organization: WHO*, Sep 2020.

Appendices

A Factors and variables for stunting

#	Factors	Corresponding variable
1	Childs age	c_age
2	Sex of child	c_sex
3	Birth weight	-
4	Birth size	-
5	Mothers health	-
6	Mothers education	mo_noedu, mo_primary, mo_secondary
7	Mothers age	mo_age_birth
8	Mothers BMI	-
9	Breastfeeding	mo_breastfeeding
10	Wealth of household	Y_his
11	Social inequality	income_quintile
12	Source of drinking water	water_improved_total
13	Sanitation	sani_improved_total
14	Place of residence	urban

Table 3: Factors associated with childhood stunting, wasting and underweight. Sources: [2] [7]

B Factors and variables for under-five mortality

#	Factors	Corresponding variable
1	Sex of child	c_sex
2	Birth order	c_first
3	Malnutrition	-
4	Vaccinations	-
5	Access to postnatal care	mo_assistance
6	Mothers age	mo_age_birth
7	Education	mo_noedu, mo_primary, mo_secondary
8	Breastfeeding	mo_breastfeeding
9	Place of delivery	-
10	Household wealth	Y_his
11	Place of residence	urban
12	Source of drinking water	water_improved_total
13	Sanitation	sani_improved_total

Table 4: Factors associated with under-five mortality. Sources: [4][12] [5]

C Correlation-matrix for stunting

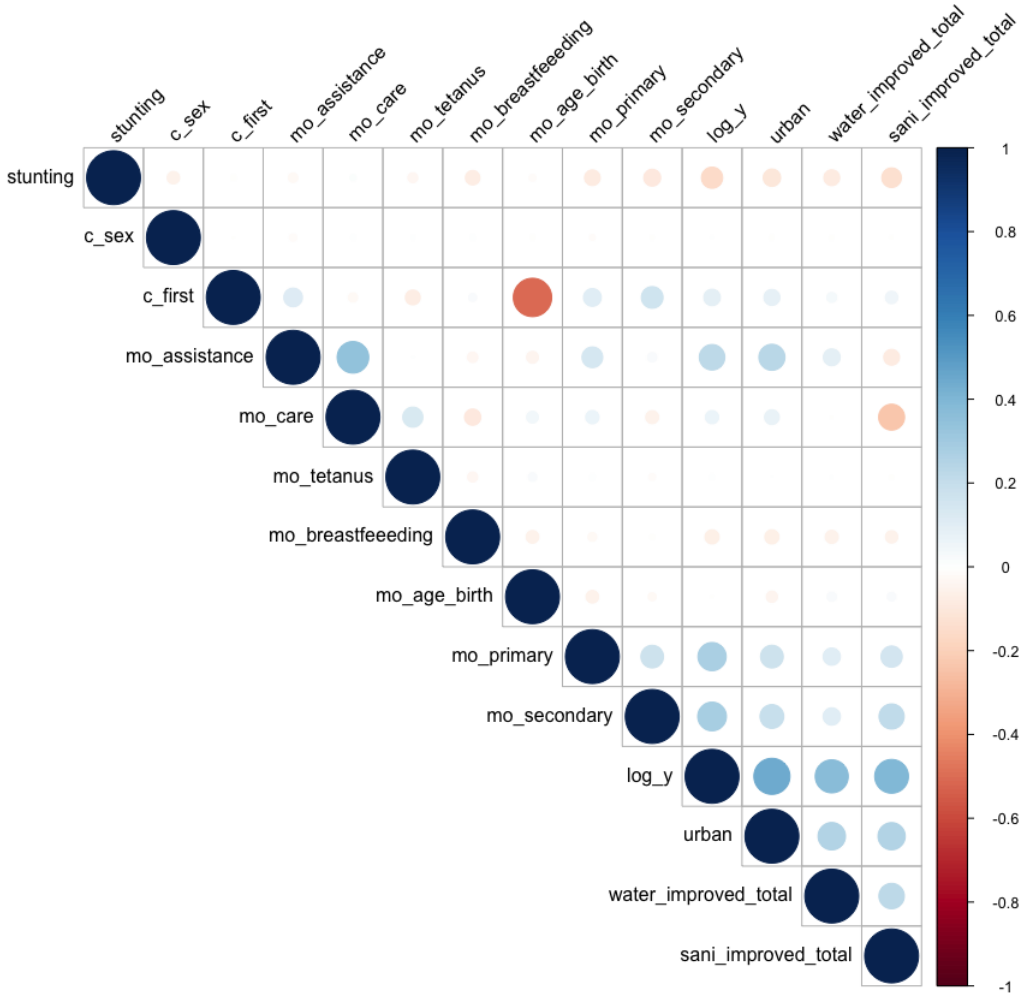


Figure 5: Correlation-matrix that should help to find variables that have a significant effect on stunting.

D Correlation-matrix for under-five mortality

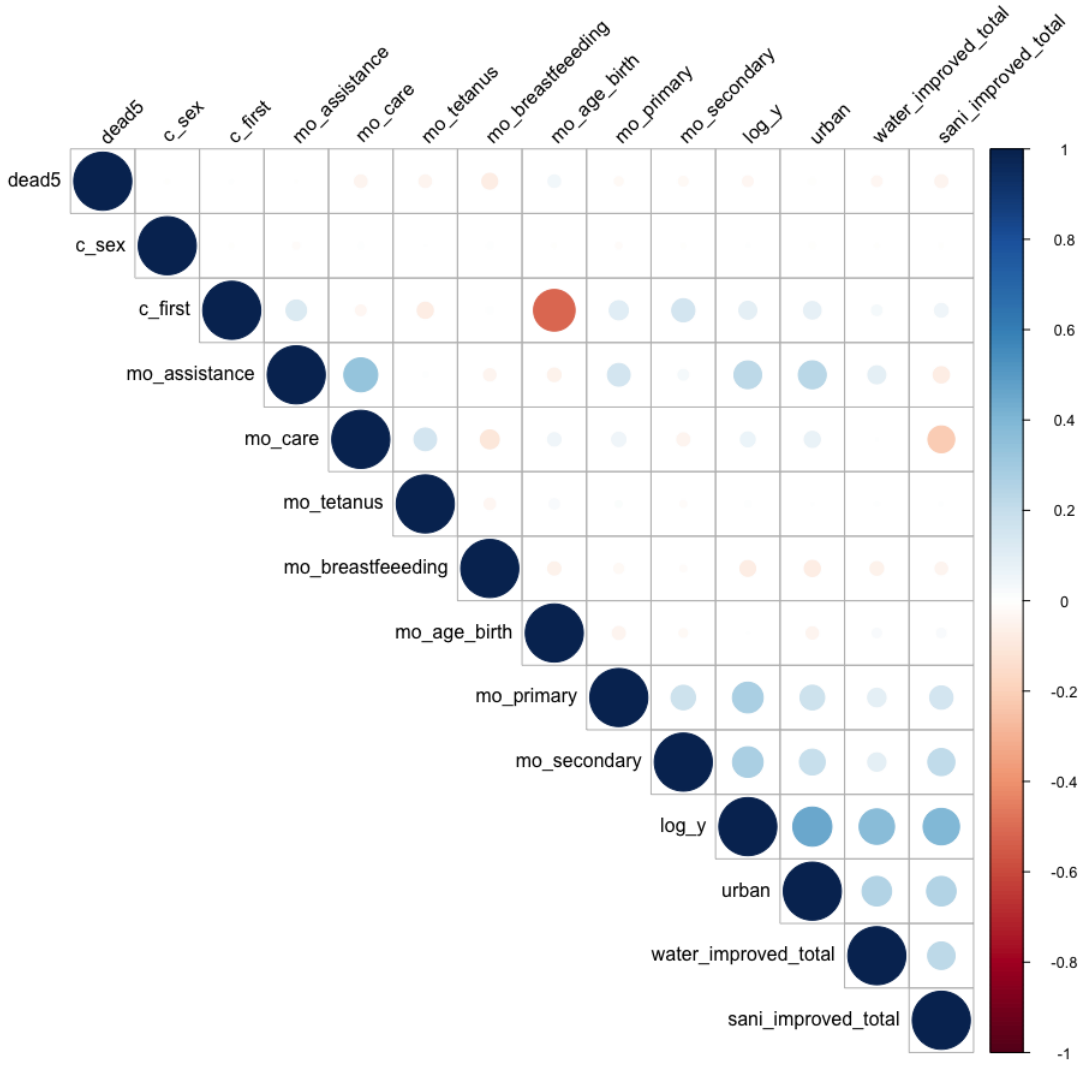


Figure 6: Correlation-matrix that should help to find variables that have a significant effect on under-five mortality.