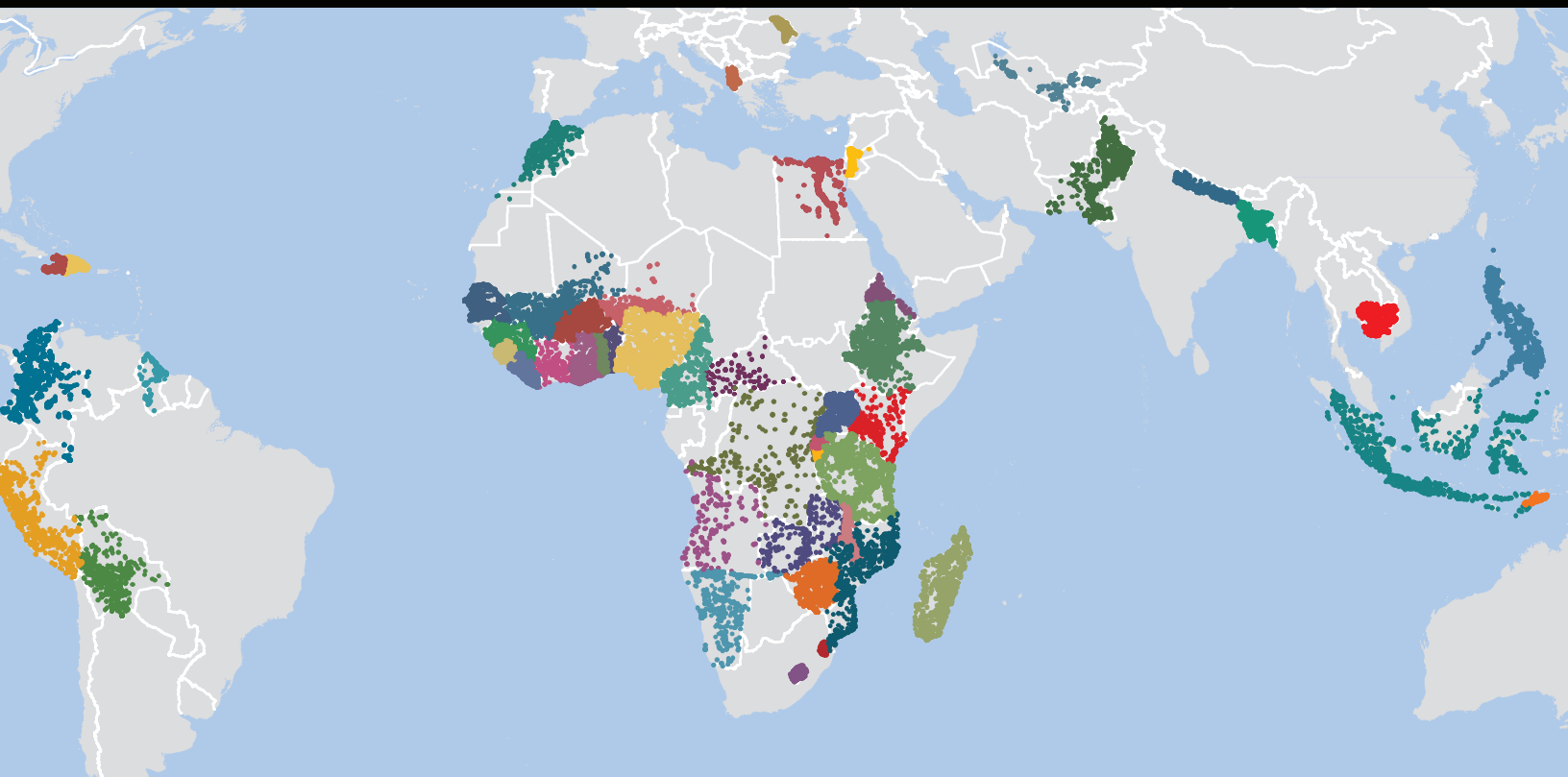




USAID
FROM THE AMERICAN PEOPLE

A PRIMER ON THE DEMOGRAPHIC AND HEALTH SURVEYS PROGRAM SPATIAL COVARIATE DATA AND THEIR APPLICATIONS

DHS SPATIAL ANALYSIS REPORTS 16



July 2018

This publication was produced for review by the United States Agency for International Development (USAID). The report was prepared by Trinadh Dontamsetti, Shireen Assaf, Jennifer Yourkavitch, and Benjamin Mayala.

DHS Spatial Analysis Reports No. 16

**A Primer on The Demographic and Health Surveys Program
Spatial Covariate Data and Their Applications**

Trinadh Dontamsetti
Shireen Assaf
Jennifer Yourkavitch
Benjamin Mayala

ICF
Rockville, Maryland, USA

July 2018

Corresponding author: Trinadh Dontamsetti, International Health and Development, ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA; phone: 301-572-0562; email: trinadh.dontamsetti@icf.com

Acknowledgments: The authors wish to thank Clara Burgert-Brucker and Trevor Croft for their comments on an earlier version of the report. The authors also wish to thank Tom Pullum, Lindsay Mallick, Rukundo Benedict, and Tom Fish for their assistance and invaluable guidance in developing this report.

Editor: Bryant Robey

Document Production: Joan Wardell

This study was carried out with support provided by the United States Agency for International Development (USAID) through The DHS Program (#AID-OAA-C-13-00095). The views expressed are those of the authors and do not necessarily reflect the views of USAID or the United States Government.

The DHS Program assists countries worldwide in the collection and use of data to monitor and evaluate population, health, and nutrition programs. Additional information about The Demographic and Health Surveys Program can be obtained from ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850 USA; telephone: +1 301-407-6500, fax: +1 301-407-6501, email: info@DHSprogram.com, internet: www.DHSprogram.com.

Recommended citation:

Dontamsetti, Trinadh, Shireen Assaf, Jennifer Yourkavitch, and Benjamin Mayala. 2018. *A Primer on The Demographic and Health Surveys Program Spatial Covariate Data and Their Applications*. DHS Spatial Analysis Reports No. 16. Rockville, Maryland, USA: ICF.

CONTENTS

TABLES	v
FIGURES	v
PREFACE	vii
ABSTRACT	ix
ACRONYMS AND ABBREVIATIONS	xi
1 BACKGROUND OF THE REPORT	1
1.1 DHS Spatial Covariates	1
1.1.1 Curation of spatial covariates.....	2
1.1.2 Extraction of spatial covariates	4
2 DATA AND METHODS	7
2.1 Study Area and Description	7
2.2 Outcomes.....	8
2.3 Sociodemographic and Spatial Covariates	9
2.4 Analytical Methods	10
3 CASE STUDIES	13
3.1 Anemia in Children Age 6-59 Months	13
3.2 Basic Vaccination Coverage of Children Age 12-23 Months	13
4 RESULTS	15
4.1 Results from Correlation Matrix	15
4.2 Results of Logit Regression–Case Study 1: Anemia in Children Age 6-59 Months	16
4.3 Results of Logit Regression–Case Study 2: Full Vaccination Coverage for Children Age 12-23 Months.....	19
5 DISCUSSION	21
5.1 Evaluation of Spatial Covariates	21
5.3 Suggestions for Further Study	22
APPENDICES	23
APPENDIX A DOCUMENTATION FOR INCLUDED COVARIATES	25
APPENDIX B SUBSET OF RWANDA 2014-15 DHS SPATIAL COVARIATE DATASET	31
APPENDIX C ANEMIA MODEL SELECTION–FULL MODEL	33
APPENDIX D VACCINATION MODEL SELECTION–FULL MODEL	35
REFERENCES	37

TABLES

Table 1	Year one spatial covariates.....	3
Table 2	Characteristics of DHS surveys	8
Table 3	Anemia model selection.....	17
Table 4	Vaccination model selection	19
Appendix B	Subset of Rwanda 2014-15 DHS spatial covariate dataset.....	31
Appendix Table C	Anemia model selection–Full model for each survey	33
Appendix Table D	Vaccination model selection–Full model for each survey.....	35

FIGURES

Figure 1	Spatial covariate rasters	4
Figure 2	Buffered raster extraction methodology	5
Figure 3	Map of study countries	7
Figure 4	Percentage of children age 6-59 months with anemia and percentage of children age 12-23 months who have all basic vaccinations	8
Figure 5	Analytical approach.....	11
Figure 6	Correlation matrix covariates using Rwanda survey data	16

PREFACE

The Demographic and Health Surveys (DHS) Program is one of the principal sources of international data on fertility, family planning, maternal and child health, nutrition, mortality, environmental health, HIV/AIDS, malaria, and provision of health services.

The DHS Spatial Analysis Reports supplement the other series of DHS reports to meet the increasing interest in a spatial perspective on demographic and health data. The principal objectives of all DHS report series are to provide information for policy formulation at the international level and to examine individual country results in an international context.

The topics in this series are selected by The Demographic and Health Surveys Program in consultation with the U.S. Agency for International Development. A range of methodologies are used, including geostatistical and multivariate statistical techniques.

It is hoped that the DHS Spatial Analysis Reports series will be useful to researchers, policymakers, and survey specialists, particularly those engaged in work in low- and middle-income countries, and will be used to enhance the quality and analysis of survey data.

Sunita Kishor
Director, The DHS Program

ABSTRACT

The Demographic and Health Surveys Program geospatial team has prepared a set of geospatial covariates datasets that provide researchers and analysts with powerful demographic, environmental, and geophysical variables to include in their analysis of DHS survey data. Often, individuals attempt to manually source and link these variables to DHS data, but are unfamiliar with Geographic Information Systems (GIS) data sources or lack the necessary skills to correctly extract and perform this link. The DHS Program has therefore prepared these files with an intuitive method of linking them to DHS survey data, allowing individuals of any statistical or geospatial analytical skill to incorporate geospatial data into their work. The DHS Program will continually update newly released surveys with the current suite of covariates and will seek to develop and release new covariates annually. This report examines methods of preparing the data for fitting a regression model using these covariates. Through two case studies, we examine the model-building process required to study the effects of spatial factors on the prevalence of anemia in children and the receipt of all basic vaccinations for children.

KEY WORDS: anemia, vaccination, spatial variables, geographic information system, model building

ACRONYMS AND ABBREVIATIONS

AIC	Akaike's information criterion
AIS	AIDS Indicator Survey
BCG	Bacille Calmette-Guerin
DHS	Demographic and Health Survey
DPT	Diphtheria-Pertussis-Tetanus
EA	enumeration area
EVI	enhanced vegetation index
GIS	geographic information system
GPS	global positioning system
GSL	growing season length
MAP	Malaria Atlas Project
MIS	Malaria Indicator Survey
SAR	Spatial Analysis Report
SES	socioeconomic status

1 BACKGROUND OF THE REPORT

The Demographic and Health Surveys (DHS) Program provides a unified set of survey and geospatial data to conduct analysis and inform policy. In recent years, however, it became apparent that many users have attempted to or expressed interest in linking DHS data to spatial data. Since September 2017, the DHS Program has been providing a set of geospatial covariates in addition to survey cluster Global Positioning System (GPS) data collected during each survey. These covariate datasets are available for download in two places: the DHS Program website (<https://dhsprogram.com/data/available-datasets.cfm>) and the Spatial Data Repository (<https://spatialdata.dhsprogram.com/covariates/>). The covariate datasets are produced using both DHS GPS data and publicly available external datasets through a standardized extraction method. While datasets including these covariates are available elsewhere online, linking these data to DHS survey data is often problematic. The DHS geospatial team sought to simplify the process of conducting geospatial analyses for experienced data users and provide a user-friendly introduction to these analyses for inexperienced users by preparing and releasing the covariate datasets. Following their release, a DHS blog was written (<https://blog.dhsprogram.com/spatial-covariates/>) announcing availability of the datasets and offering potential users information on the covariates, but further guidance on the practical usage of these data was necessary.

Accordingly, this report will provide the information that users need to understand the geospatial covariate datasets and to begin using them for their own analyses. This document does not provide a comprehensive review of the development of these datasets, which is described in a PDF file included with every covariate dataset. (Appendix A provides examples of these files for the covariates used in this report.) Rather, this report presents guidance in the form of practical applications of analysis including geospatial covariates.

This Spatial Analysis Report (SAR) is intended for all users conducting statistical analysis of DHS survey data, irrespective of their skills or knowledge of geospatial analysis. The authors hope that all users find this a useful primer for incorporating geospatial covariates in analyses of DHS survey data, especially those who have not previously attempted to conduct such analysis due to unfamiliarity with Geographic Information Systems. The authors also hope this document will give these users confidence in their ability to bring a spatial dimension to their work. This report includes two main sections: a description of the covariate data—including the efforts to determine an appropriate set of covariates, extraction methods, and the creation of the dataset—and two case studies using two health outcomes as examples of including geospatial covariates in analysis.

1.1 DHS Spatial Covariates

The DHS spatial covariates came from various data sources and were not consistently or easily linked to DHS data. Further, many users were unfamiliar with the GIS skills required to perform these linkages and were discouraged from further incorporating spatial elements in their work. To address these issues, the DHS Program prepared a standardized set of publicly available geospatial covariates that include many commonly used variables. These covariates can be linked easily to DHS survey data without the need for GIS software, allowing researchers who may lack GIS experience to include a spatial component in their analysis.

1.1.1 Curation of spatial covariates

The DHS Program conducted interviews with geospatial experts to obtain guidance for curating the list of covariates. Seventeen individuals and groups provided feedback that focused on:

- What they thought DHS should include as possible covariates,
- What specific datasets they recommend for various topics (such as for rainfall or temperature),
- What specific concerns they had regarding use of these data.

The DHS Program also disseminated a short survey to all DHS dataset users who had downloaded any type of data from the DHS Program website between April 2016 and March 2017, seeking feedback on their topic areas of interest, their interest in using geospatial data in their work, and their current level of experience with geospatial data. Finally, the DHS Program conducted a literature review of publications between 2001 and 2016 that used DHS data. As a result of this work, 192 geospatial covariates were identified and divided into eight categories: agriculture, climate, environment, health condition, infrastructure, physical earth, political, and population.

The DHS Program identified geospatial covariates to be prepared by combining the results of the literature review with the guidance provided by the respondents of the user survey and the expert interviews to select external datasets based on four key criteria:

- Datasets must have either global or regional extent, such as coverage for all of Africa or Asia. As such, data organized by a single country or not available for the majority of countries were not considered.
- Datasets must be publicly available.
- Datasets must have a well-documented acquisition or creation process along with detailed metadata of both inputs and processing procedures.
- Datasets must have data available for relevant timeframe(s), determined during the preparation phase.

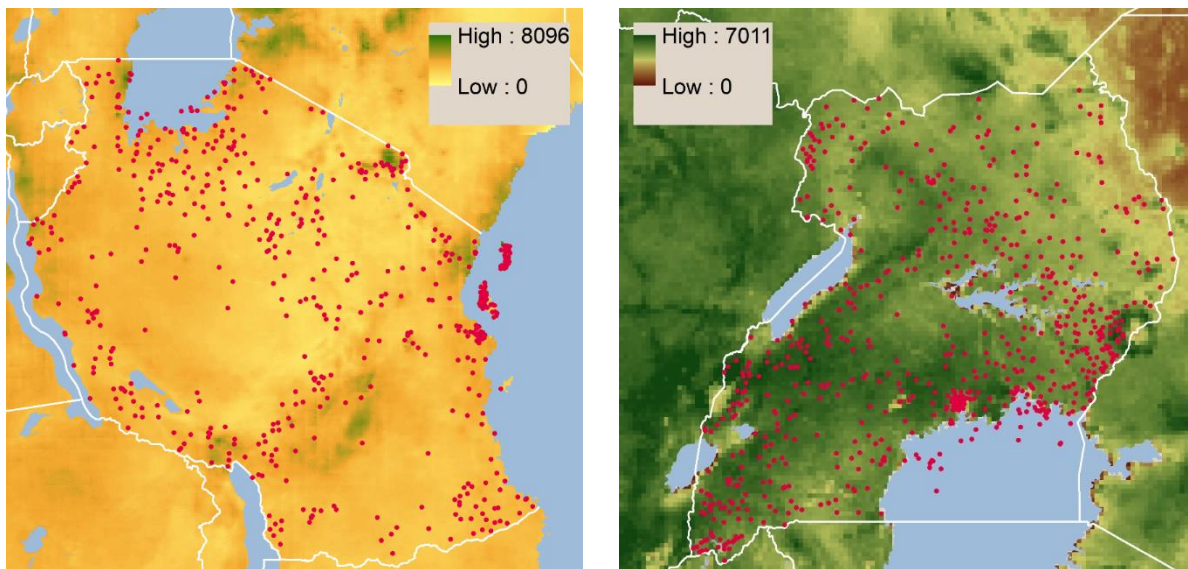
These selection criteria helped the DHS Program to curate the list of covariates to be prepared, resulting in a set of covariates that are highly standardized. A total of 22 covariates were prepared from 15 dataset sources. More information regarding each covariate's specific data source, unit, timeframe, data description, and full citation can be found in the data definition file included with every covariate dataset download, and reproduced in Appendix A. A full list of covariates prepared during the first year of this activity can be found in Table 1.

Table 1 Year one spatial covariates

Topic	Covariate	Timeframe	Units	Data Source
Agriculture	Drought Episodes	Single Dataset	Count of Episodes, 1980-2000	SEDAC
Agriculture	Growing Season Length [§]	Single Dataset	Mode of 16 Categories Representing a Range of the Number of Days Within the Period of Temperatures Above 5C When Moisture Conditions Are Considered Adequate	IIASA/FAO
Environment	Aridity	Single Dataset	Unitless Index. Higher Aridity Index (AI) Reflect Greater Humidity. Lower AI Reflects Greater Aridity.	CGIAR
Environment	Enhanced Vegetation Index [§]	1985-2015 in 5-yr Increments	Unitless Index. Higher Scores Indicate Higher Vegetation Vigor/Photosynthetic Activity	AVHRR (1985-1995), MODIS (2000-2015)
Environment	Potential Evapotranspiration	Single Dataset	Monthly Estimate of Pet (Mm/Month) Averaged Over the 1950-2000 Period	CGIAR
Environment	Proximity to Coast/ Large Lakes [§]	N/A	Meters	GSHHG
Environment	Proximity to Protected Areas	N/A	Meters	Protected Planet
Environment	Rainfall [§]	1985-2015 in 5-yr Increments	Mm/Year	CHIRPS
Environment	Slope	Single Dataset	Meters Above Sea Level	SRTM
Environment	Temperature (Average Monthly)	January - December Averages for 1970-2000	Degrees Celsius	WorldClim
Health	ITN Net Coverage	2000-2015 in 5-yr Increments	Percent of People Who Slept under an Insecticide-Treated Net (ITN) on Any Given Night	MAP
Health	Malaria	2000-2015 in 5-yr Increments	<i>Plasmodium falciparum</i> Malaria Cases Per Person Per Year Observed	MAP
Infrastructure	Nightlights [§]	2015 Annual Composite	Average Cloud-Free Radiance Values	VIIRS (DNB)
Infrastructure	Proximity to National Border	2014	Meters	State Dept. LSIB
Infrastructure	Travel Times	Year 2000 Estimate	Estimated Travel Time (Minutes) to the Nearest City of 50,000 or More People in Year 2000	FOROBS
Infrastructure	Urbanization - Global Human Footprint	1995-2004 (Overall)	Unitless Index. Higher Values Indicate Higher Human Influence.	SEDAC
Infrastructure	Urbanization - GHS Settlement Grid	1990, 2000, 2015	Mode of Categories: 1 = "Rural Cells" or Base (BAS); 2 = "Urban Clusters" or Low-Density Clusters (LDC); 3 = "Urban Centers" or High-Density Clusters (HDC)	GHSL-SMOD (EC JRC)
Infrastructure	Urbanization - GHS Build-Up Grid	1990, 2000, 2015	Built-Up Presence Index, Range 0-1. Higher Value Is More Confidence that It Is Built Up	GHSL (Landsat) (EC JRC)
Population	UN Adjusted Population Count	2000-2015 in 5-yr Increments (Adjusted)	Number of People	GPW v4
Population	UN Adjusted Population Density	2000-2015 in 5-yr Increments (Adjusted)	Number of People/Km ²	GPW v4
Population	Count [§]	2000-2015 in 5-yr Increments (UN Adjusted)	Number of People	WorldPop
Population	Density	2000-2015 in 5-yr Increments	Number of People/Km ²	WorldPop

Note: § indicates the covariates that were used in this report.

Figure 1 Spatial covariate rasters



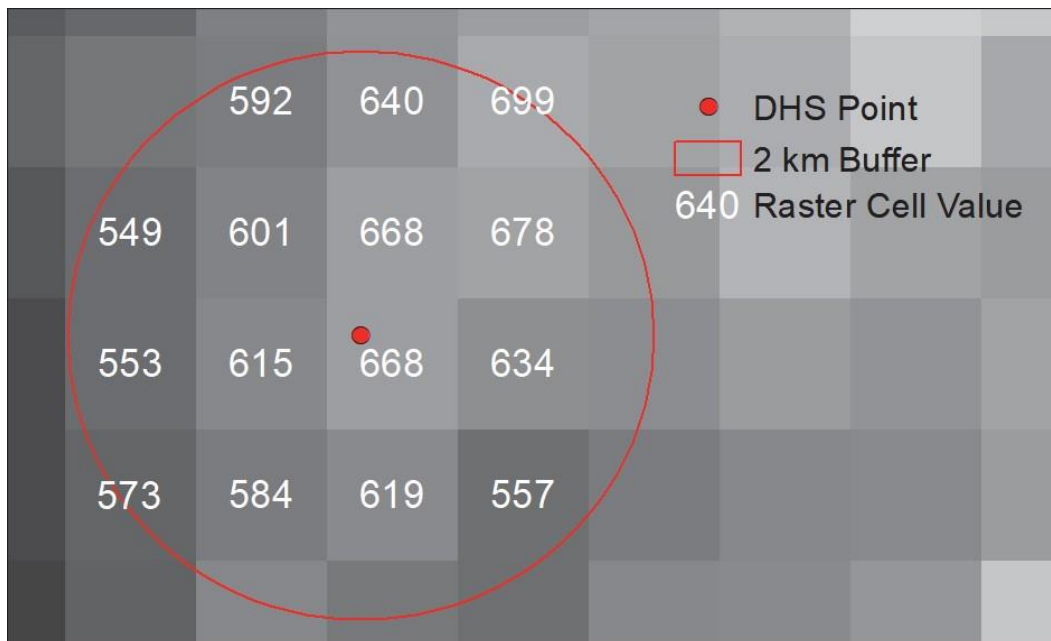
For illustrative purposes, Figure 1 shows the GPS data points for the Tanzania 2015-16 DHS overlaid atop the rainfall covariate raster dataset (left), and the Uganda 2016 DHS points atop the Enhanced Vegetation Index (EVI) covariate raster (right). The GPS data are used to extract the covariate data from the source dataset, as explained in the following section.

1.1.2 Extraction of spatial covariates

These geospatial covariates were grouped into two categories—neighborhood calculations and distance calculations—based on the nature of the covariate and the corresponding extraction method it necessitated.

For neighborhood calculations, the covariate data are extracted based on a raster extraction method specified in *Spatial Analysis Report 8* (Perez-Heydrich et al. 2013) and further discussed in other literature (Warren et al. 2016). Briefly, a circular buffer with a 10 kilometer radius is drawn around rural points, while a 2 kilometer buffer is drawn around urban points. All raster cells whose centroids fall within the buffer surrounding a point are used in the extraction calculation, while raster cells whose centroids do not fall within the buffer are excluded. The appropriate calculation (mode, mean, etc.) for each individual covariate is performed using the *Zonal Statistics as Table* tool through ArcPy. This calculates the chosen metric for the covariate for each cluster in the input DHS dataset, and these results are exported to a table. Figure 2 gives an example of this methodology.

Figure 2 Buffered raster extraction methodology



For distance calculations, vector data are used rather than rasters. Using the *Near* tool in ArcMap, which calculates the geodesic distance between each DHS point and the nearest boundary of a selected polygon, proximity covariates were prepared. These included distance of the DHS cluster to the nearest national border, to coasts and large lakes, and to protected areas. These results are appended to the same table within which the results of the neighborhood calculation are exported for each point. An example of the exported data can be found in Appendix B, which shows the data of the chosen spatial covariates for the first 30 clusters of the Rwanda 2014-15 DHS survey.

2 DATA AND METHODS

2.1 Study Area and Description

The DHS Program is a leader in collecting and providing survey data on core development indicators. Most surveys now provide geocoded data for individual survey clusters (enumeration areas (EAs)). GPS coordinates for DHS survey clusters provide a method of linking local-scale data to survey outputs for analysis of demographic and health status indicators.

This analysis uses data from six surveys in four neighboring East African countries—Burundi 2017 DHS, Rwanda 2014-15 DHS, Uganda 2016 DHS, Tanzania 2008 THMIS (a combined AIS/MIS survey), 2010 DHS, and 2015 DHS—to explore the process of building a regression model using the spatial covariate datasets. These models are then used to study the impact of geospatial factors on the prevalence of anemia in children age 6-59 months and receipt of all eight basic vaccinations for children age 12-23 months. Any anemia in children is defined by hemoglobin levels of <11.0 g/dL. Receipt of all eight basic vaccinations is considered to include bacille Calmette-Guérin (BCG), three doses of Diphtheria-Pertussis-Tetanus (DPT) or a DPT-containing vaccination such as a Pentavalent vaccine, three doses of oral polio vaccine, and one dose of a measles-containing vaccine.

Figure 3 Map of study countries



Figure 3 shows the four countries with DHS survey data used for this analysis.

Table 2 Characteristics of DHS surveys

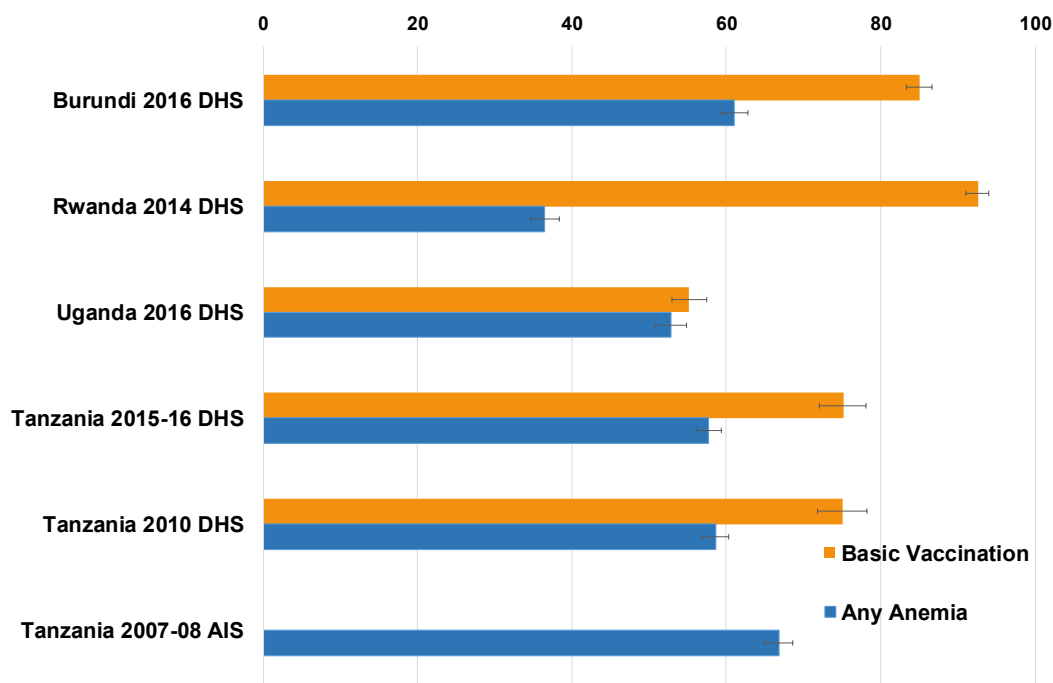
DHS Survey	Number of Households Interviewed	GPS Clusters Collected
Burundi 2016 DHS	16,620	554
Rwanda 2014 DHS	12,793	492
Uganda 2016 DHS	20,880	696
Tanzania 2015-16 DHS	13,376	608
Tanzania 2010 DHS	10,300	475
Tanzania 2007-08 AIS	9,144	475

Table 2 describes characteristics of the six DHS surveys involved in this analysis, specifically the number of households interviewed during fieldwork for each survey and the corresponding number of survey clusters for which GPS data were collected.

2.2 Outcomes

The outcomes for the analysis in this report are split into two case studies: the first concerns any anemia in children age 6-59 months, and the second, receipt of all basic vaccinations for children age 12-23 months. As Figure 4 shows, the burden of anemia in children is high across the countries of interest. Burundi, Uganda, and Tanzania all have prevalence over 50%, while about a third of children in Rwanda have any anemia. Burundi and Rwanda fare best in vaccination coverage: both countries have greater than 80% coverage. However, both Uganda and Tanzania are below 80%. Section 3 provides the background on these two case studies.

Figure 4 Percentage of children age 6-59 months with anemia and percentage of children age 12-23 months who have all basic vaccinations



2.3 Sociodemographic and Spatial Covariates

A review of the literature on anemia prevalence and vaccination coverage revealed two things about these studies: (1) most focused on the nonspatial factors affecting anemia and vaccination, and (2) those that did venture into spatial analysis of anemia or vaccine coverage either stratified their data by urban or rural settings (Kuziga, Adoke, and Wanyenze 2017; Canavan et al. 2014), or assessed the spatial relationships of the survey data by evaluating spatial autocorrelation (Ejigu, Wencheke, and Berhane 2018). However, we did not find widespread use of spatial covariates such as rainfall, vegetation, or population estimates throughout the literature, which we believe indicates a lack of accessibility to these data and provides further evidence of the need for such data to be readily available and easily accessed. This gave us the opportunity to explore the impact that these environmental and geophysical factors potentially have on both the prevalence of anemia and the receipt of basic vaccinations.

Rainfall creates conditions that allow sufficient surface water for mosquito breeding, and is often studied for its effects on transmission of malaria (Arab, Jackson, and Kongoli 2014; Sewe et al. 2015). However, research on direct interactions between rainfall and anemia is limited. One study did take droughts into consideration in its analysis, but primarily in response to conditions in the field during data collection for a malaria prevention trial (Gari et al. 2017). Given the association between rainfall and malaria transmission, along with rainfall's role in promoting vegetation growth, we included the DHS Program's rainfall covariate, defined as the average rainfall of the cells whose centroid falls within a radius of 10 km (for rural points) or 2 km (for urban points).

Similarly to rainfall, vegetation cover linked to population data has been identified as a predictor for malarial transmission. Studies have shown the risk of dense vegetation cover near households as a facilitator of malaria transmission (Ricotta et al. 2014). For instance, farmers clearing vegetation to gain more productive land can often inadvertently lead to the creation of breeding sites for mosquitoes (Walsh, Molyneux, and Birley 1993). Increased agricultural development and irrigation, such as rice farming, can increase the transmission of malaria due to creation of large stagnant pools of water that provide breeding habitats favorable to *Anopheles gambiae* and, in Tanzania, *Anopheles arabiensis* larvae (Janko et al. 2018; Mboera et al. 2010; Ernst et al. 2009). Given their indirect associations with anemia through malaria transmission and the relative paucity of literature examining them in the context of anemia, we felt it important to include the Enhanced Vegetation Index covariate along with the Growing Season Length (GSL) covariate as part of our analysis. Malaria was not used as a covariate in the anemia models. This was because the malaria covariate is produced by the Malaria Atlas Project (MAP), which uses some of the spatial covariates in their predictive models (for instance, nighttime lights). These spatial covariates are also used in our models and therefore would have produced some bias. Definitions for these covariates can be found in Appendix A.

The literature also helped us choose typical sociodemographic controls for the analysis of both our case studies. In sub-Saharan Africa, for example, sex of child has been studied extensively in this context (Ncogo et al. 2017; Ngesa and Mwambi 2014). In addition, child's age, categorized as 5-29 months and as 30-59 months, was included as a control. In our analysis, age category was not included in the vaccination model, which is only relevant for children age 12-23 months. Caretaker's education level is also consistently included as a control variable (Simbauranga et al. 2015; Ismael, Olivier, and Barthelemy 2007). In our model, education was categorized as none, primary, and secondary or above. Because studies often measure

outcomes by socioeconomic status (SES) (Foote et al. 2013), our analysis used the DHS Program's standard household wealth index variable (categorized into quintiles: lowest, second, middle, fourth, and highest) as a covariate.

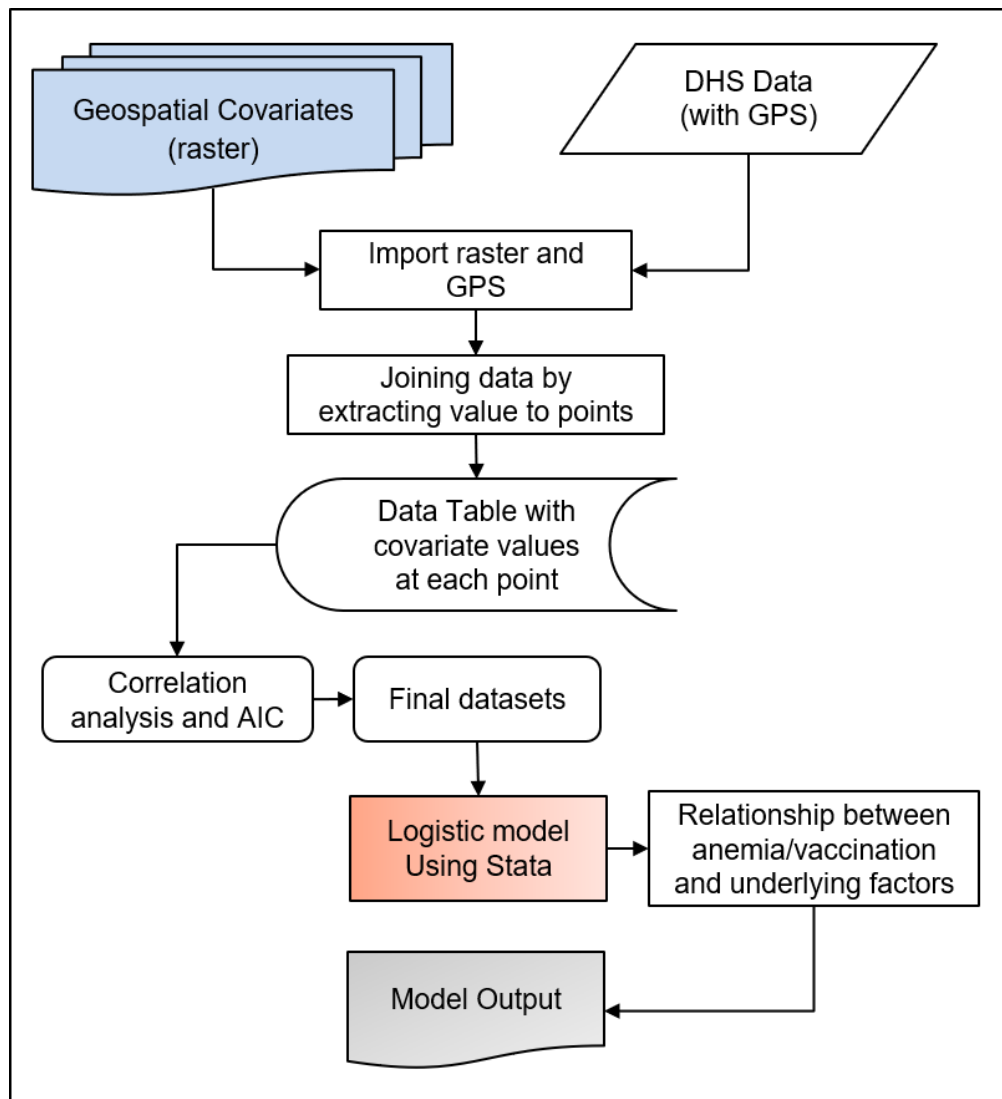
Other geospatial covariates, such as nightlights, UN population count, and proximity to coast/large lakes, were selected for inclusion in our analysis specifically because of the relative lack of literature that explores the impact of these factors on either anemia prevalence or vaccination coverage. We were unable to find any such research. We believe this report is an opportunity to conduct an exploratory analysis to examine the potential links between these covariates and the two outcomes.

2.4 Analytical Methods

This analysis uses Stata 14 software to conduct a simple logit regression model of anemia in children age 6-59 months and basic vaccination for children age 12-23 months. This is accomplished by using DHS survey data for each of the six surveys included, along with each survey's corresponding geospatial covariate data. For each survey, correlation matrices were constructed for the covariates to be included in the model in order to identify potential confounding as a result of highly correlated independent variables. A covariate was determined to be too highly correlated to another if the absolute value of the correlation coefficient was greater than or equal to 0.7. These matrices can be found in Section 4.1 of this report.

A regression model was fit for each spatial covariate separately for each outcome along with the control variables. After examining the significance of the geospatial covariate, a final model was fit for each survey and each outcome that included the controls and all the significant geospatial covariates. The significant geospatial variables were only added together in the model if they were not highly correlated. In addition to examining the significance of the covariates, we computed the Akaike Information Criterion (AIC) for the models and compared it with the AIC of the base model, which includes only the control variables. A substantial reduction in the AIC from the base model would indicate that is important to include the geospatial covariate in the model. Selecting the geospatial covariate to include in the final model based on its significance is generally in agreement with selecting it based on the reduction in the AIC value. If two covariates that are highly correlated are both significant, then the covariate that provides a model with a lower AIC value is selected for inclusion. AIC were computed considering the sampling weight but not the stratification design of the survey. This was because the STATA command used to produce the AIC (*estat ic*) does not support the use of survey design commands. For the logit regression models, the survey stratification and sampling weights were accounted for. Figure 5 presents a summary of our approach.

Figure 5 Analytical approach



Flow diagram showing the steps used for modeling the relationship between anemia/vaccination and the underlying factors

In addition to the comparison of the logit regression model-building procedures for the two outcomes across all surveys, one country was selected to demonstrate the model-building procedure for the two outcomes across time. Tanzania was selected because it has three surveys since 2000 with GPS data (2007-08 THMIS, 2010 DHS, and 2015-16 DHS). This analysis was performed to demonstrate the possible change in the inclusion of the geospatial covariates for the same country across time. As there could be a lag between the measure of the covariates and their impacts on survey respondents, each covariate was selected for the most recent year *preceding* the survey for which data were available.

3 CASE STUDIES

3.1 Anemia in Children Age 6-59 Months

Anemia is a condition characterized by low hemoglobin concentration in the blood, and it has significant health, social, and economic consequences. It is associated with poor birth outcomes, including low birth weight, prematurity, and perinatal and maternal mortality (Balarajan et al. 2011). It is also associated with poor cognitive function and motor development in children (WHO 2017b). In addition, the collective burden of anemia has been estimated to account for decreased economic growth and millions of years lived with disability in 2010 alone (WHO 2017b). Although the prevalence of anemia decreased between 1990 and 2010, it is estimated to affect one-third of the world's population, including 800 million women and children (WHO 2017b).

Causes of anemia are generally classified in terms of nutrient deficiencies, infections and disease, and genetic disorders. Of the main three causes of anemia globally—iron deficiency, hemoglobinopathies, and malaria—iron deficiency contributes to 42% of anemia cases among children under age 5 (WHO 2017b). Children under age 5 have the largest burden of anemia globally because of their rapid growth and high iron requirements, especially in the first two years of life (WHO 2017b). Poor iron status can be transferred from mother to child (WHO 2017b; Khan, Awan, and Misu 2016), and typical complementary foods are low in iron (WHO 2017b).

While in adults there is an increased risk of anemia for women of reproductive age, male children are more likely to be anemic than female children (Legason et al. 2017; Ngesa and Mwambi 2014). Infant iron stores are lower in males than in females during the first year of life (WHO 2017b). Gendered feeding practices may also contribute to differences in anemia between boys and girls. Urban-rural location, household wealth status, and maternal education are also associated with children's anemia status (WHO 2017b).

There is also regional variation in anemia prevalence, depending in part on geospatial factors such as malaria prevalence and coverage of insecticide-treated bed nets (ITNs). In Tanzania, malaria is associated with severe anemia in children (Simbauranga et al. 2015). Regional variation signals the importance of investigating the contribution of geospatial covariates to the burden of anemia. The purpose of this case study is to illustrate how geospatial covariates can be used with demographic covariates to analyze the contribution to the burden of anemia in children under age 5.

3.2 Basic Vaccination Coverage of Children Age 12-23 Months

Immunization is one of the most successful, equitable, and cost-effective interventions, preventing 2–3 million deaths each year (WHO 2017a), and basic vaccination coverage is a measure of both child health and health system performance. An estimated 19.5 million infants are still missing basic vaccine coverage (WHO 2018); in 2016, one in every 10 infants around the world did not receive any vaccinations (WHO 2017a). In addition to measles vaccine coverage, the addition of new vaccines (pneumococcal and rotavirus) to the list of recommended routine child immunizations holds promise for reducing the prevalence of child deaths due to pneumonia and diarrheal disease, respectively (Liu et al. 2015). The World Health Organization (WHO) and partners developed the Global Vaccine Action Plan to guide efforts to ensure equitable access to vaccines, but progress toward its 2020 targets is not on track (WHO 2018).

DHS Spatial Analysis Report 12 (Burgert-Brucker et al. 2015) identified uneven distribution of immunization coverage between and within 27 countries in sub-Saharan Africa. This varying immunization coverage within countries has led to recommendations to address regional variation with different strategies for delivery of information and services (Lakew, Bekele, and Biadgilign 2015).

4 RESULTS

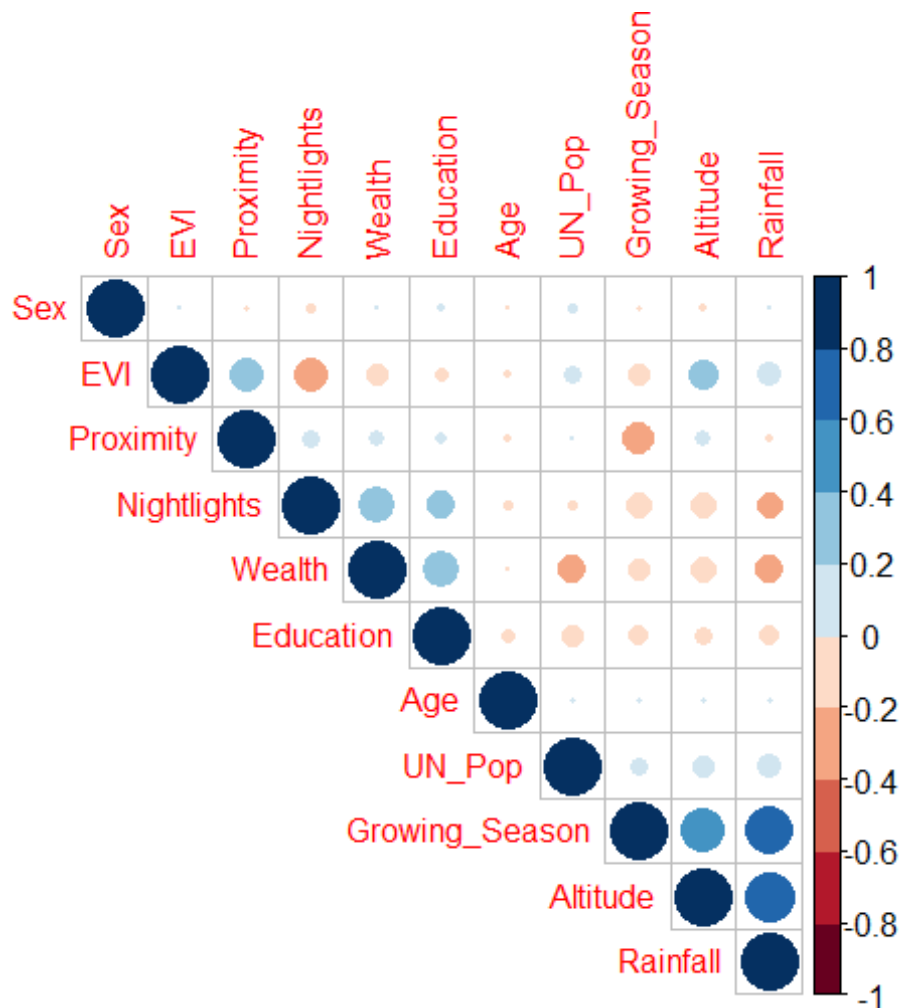
4.1 Results from Correlation Matrix

Before running the final logit models for each outcome, we examined the results of the correlation testing for the independent variables, which helped us to determine, between a pair of highly correlated, statistically significant covariates, which variables to select for inclusion in the models. In the Tanzania 2008 survey, both the proximity and elevation covariates were statistically significant per the model-building exercise (proximity: $p < .001$, elevation: $p < .001$), but were highly correlated per the correlation matrix (correlation coefficient of 0.7967). Reviewing the relative change to AIC attributed to each covariate (proximity: -32.5 relative to base model; elevation: -266.7 relative to base model), we decided to include the elevation covariate in our final logit regression model rather than the proximity covariate, because it contributed a larger change in the model AIC.

Similarly, the Tanzania 2010 and Tanzania 2015-16 surveys exhibited a high degree of correlation between elevation and proximity (correlation coefficient 0.8199 and 0.8005 respectively), while both covariates were shown to be statistically significant during the model building (both $p < .001$ in each survey). A review of the relative change to model AIC (Tanzania 2010: elevation: -102.8 relative to the base model, proximity: -32.7 relative to the base model; Tanzania 2015-16: elevation: -47.9 relative to the base model, proximity: -18.9 relative to the base model) helped us to determine that, as with Tanzania 2008, elevation would be included in the base model instead of proximity for both surveys due to causing a larger drop in model AIC.

Figure 6 shows an example of the correlation matrix for the covariates used in the analysis for the Rwanda survey. In the case of Rwanda, a correlation coefficient of 0.7463 was observed between the elevation (altitude) and rainfall (2010) covariates, indicating a high degree of correlation. However, the model-building exercise revealed that rainfall was not a statistically significant covariate, and as such would not be included in the final logit regression model, regardless. We therefore included elevation in the final model.

Figure 6 Correlation matrix covariates using Rwanda survey data



No other covariate pairs in any other survey exhibited high degrees of correlation that needed special attention during the logit model-building activity. As the same set of demographic and geospatial covariates was used for both case studies, the same pairs of highly correlated covariates were examined when analyzing basic vaccination among children age 12-23 months. In the vaccination analysis, no geospatial covariates were significant in the Rwanda 2014-15 DHS or any of the three Tanzania surveys, thereby rendering moot the review of relative change to model AIC. However, Burundi 2017 DHS and Uganda 2016 DHS had covariates that were significant but not strongly correlated, and so no decision had to be made regarding selecting the covariates for inclusion in the final models.

4.2 Results of Logit Regression – Case Study 1: Anemia in Children Age 6-59 Months

Table 3 summarizes the results of the model selection for the anemia outcome. The top half of the table summarizes the significance and AIC values for the models that include each geospatial covariate with the control variables. The bottom half of the table indicates the final model selected for each survey. The coefficients for the full models are summarized in Appendix C. The significance of the geospatial covariates varied by survey. However, elevation was significant for all surveys, and rainfall was significant for three of the four countries in the most recent surveys.

Table 3 Anemia model selection

Model	Burundi			Rwanda			Uganda			Tanzania 2010			Tanzania 2008		
	coef'	p-value	AIC	coef'	p-value	AIC	coef'	p-value	AIC	coef'	p-value	AIC	coef'	p-value	AIC
BM ²			6984.6			3944.5			4954.3			9558.2			7506.0
BM + 'EVI'	-0.0002	0.102	6981.4	0.0001	0.172	3944.0	-0.0002	0.075	4949.0	0.0002	0.017	9548.2	-0.0002	0.002	7480.1
BM + 'Growing'	-0.0759	0.237	6983.7	-0.1436	<0.001	3928.9	-0.0140	0.694	4956.0	0.0842	0.002	9539.2	-0.0536	0.064	7501.5
BM + 'Nightlights'	0.0316	0.733	6986.3	-0.0297	0.143	3944.3	0.0344	0.079	4952.9	0.0365	0.002	9548.5	0.0413	0.006	7495.2
BM + 'Proximity'	-0.0000007	0.662	6986.0	0.0000003	0.345	3945.4	-0.0000006	0.661	4955.9	0.0000009	<0.001	9539.2	0.0000001	<0.001	7473.3
BM + 'Rain'	-0.0009	<0.001	6939.9	-0.0005	0.086	3941.9	0.0007	<0.001	4930.6	0.0006	<0.001	9511.9	0.00005	0.772	7507.9
BM + 'UN_Pop'	-0.0001	0.858	6986.5	-0.0000003	0.666	3946.2	0.00003	0.930	4956.3	0.0000009	0.006	9554.0	0.0000001	0.864	7508.0
BM + 'Elevation'	-0.0004	0.001	6937.9	-0.0004	0.010	3935.3	-0.0010	<0.001	4911.3	-0.0004	<0.001	9510.3	-0.0006	<0.001	7403.3
<i>Full models³</i>															
BM + 'Elevation' + 'Rain'	-	-	6904.4	-	-	-	-	-	-	-	-	-	-	-	-
BM + 'Elevation' + 'Growing'	-	-	-	-	-	3929.4	-	-	-	-	-	-	-	-	-
BM + 'Elevation' + 'Rain'	-	-	-	-	-	-	-	-	4898.8	-	-	-	-	-	-
BM + All Covariates - 'Proximity' ⁴	-	-	-	-	-	-	-	-	-	-	-	9483.0	-	-	-
BM + 'EVI' + 'Nightlights' + 'Elevation' ⁴	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7365.5
BM + 'EVI' + 'Growing' + 'Elevation' + 'Rain' ⁴	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6260.3

Notes: 1 - Coefficient for the added geospatial covariate; 2 - BM=Base model. Includes the sociodemographic variables of wealth index, child's age, child's sex, and education of mother; 3 - Coefficients for the full model for each survey are found in Appendix B; 4 - Proximity omitted due to high correlation with Elevation covariate.

For Burundi, the rainfall and elevation covariates were significant and showed a substantial reduction in the AIC compared with the base model. This indicates that these variables provide a significant explanation to the model and should be included. Therefore, the final model for Burundi included both these covariates, which reduced the AIC further—from 6,984.6 for the base model to 6,904.4 in the final model. Two geospatial covariates were also significant in the Rwanda survey: growing season length and elevation. The final model that included these two covariates reduced the AIC from 3,944.5 in the base model to 3,929.4 in the final model. For Uganda, rainfall and elevation were the only two statistically significant covariates. The inclusion of these two covariates reduced the AIC from 4,954.3 in the base model to 4,898.8 in the final model.

The model selection procedure was also used for the three surveys in Tanzania. For the Tanzania 2015-16 DHS, all the geospatial covariates were significant. However, because proximity was omitted due to strong correlation with elevation, the final model for the 2015 Tanzania survey included all geospatial covariates except proximity. This reduced the AIC from 9,558.2 to 9,483.0, which indicates an improvement in the model due to the addition of the geospatial covariates. In the Tanzania 2010 DHS survey, only EVI, nightlights, proximity, and elevation were significant. In the Tanzania 2008 DHS survey, EVI, growing season length, proximity, rainfall, and elevation were all significant. Again, for the 2010 and 2008 surveys, elevation was highly correlated with proximity, and therefore proximity was excluded from the models. Across all three Tanzania surveys, the enhanced vegetation index, proximity, and elevation covariates were significant for the anemia outcome.

4.3 Results of Logit Regression – Case Study 2: Full Vaccination Coverage for Children Age 12-23 Months

Table 4 Vaccination model selection

Model	Burundi			Rwanda			Uganda			Tanzania 2015			Tanzania 2010		
	coef ¹	p-value	AIC	coef ¹	p-value	AIC	coef ¹	p-value	AIC	coef ¹	p-value	AIC	coef ¹	p-value	AIC
BM ²			2234.8			797.3			3946.6			2298.4			1678.2
BM + 'EVI'	-0.0001	0.528	2236.1	-0.0003	0.269	797.0	0.0001	0.098	3944.2	-0.00004	0.740	2300.3	0.0002	0.257	1678.0
BM + 'Growing'	-0.3103	0.005	2223.8	-0.1094	0.270	797.7	-0.1060	0.002	3936.2	-0.0477	0.375	2298.9	0.0457	0.457	1679.3
BM + 'Nightlights'	-0.1372	0.438	2235.7	-0.0157	0.753	799.2	-0.0454	0.011	3942.1	0.0658	0.066	2294.4	0.0651	0.128	1675.0
BM + 'Proximity'	0.000001	0.583	2236.4	-0.000006	0.504	798.7	0.000005	0.001	3932.1	0.0000006	0.312	2298.4	0.0000008	0.221	1677.9
BM + 'Rain'	-0.0006	0.083	2231.9	0.0008	0.272	797.0	-0.0002	0.186	3946.2	0.0001	0.586	2300.0	0.0002	0.536	1679.5
BM + 'UN_Pop'	0.0015	0.092	2231.6	0.000002	0.272	797.3	0.0006	0.078	3945.0	0.000002	0.053	2294.4	0.000001	0.421	1679.4
BM + 'Elevation'	0.0006	0.014	2218.3	-0.0003	0.544	798.6	0.0007	<0.001	3932.9	-0.0001	0.517	2299.6	-0.00002	0.887	1680.2
<i>Full models³</i>															
BM + 'Growing' + 'Elevation'	-	-	2199.8	-	-	-	-	-	-	-	-	-	-	-	-
BM + 'Growing' + 'Nightlights' + 'Elevation' - 'Proximity'	-	-	-	-	-	-	-	-	3911.2	-	-	-	-	-	-

Notes: 1 – Coefficient for the added geospatial covariate; 2 – BM=Base model. Includes the sociodemographic variables of wealth index, child's age, child's sex, and education of mother; 3 - Coefficients for the full model for each survey are found in Appendix C.

Table 4 summarizes the results of the model selection for the vaccination outcome. The coefficients for the full models are summarized in Appendix D. Very little significance in the geospatial covariates was found for the vaccination outcome. The significant geospatial covariates were only found in Burundi and Uganda. In both these countries, growing season length (GSL) was significant. For the Burundi 2017 DHS survey, elevation was also significant, and the inclusion of this covariate with the GSL covariate reduced the AIC from 2,234.8 in the base model to 2,199.8 in the final model. For Uganda 2016 DHS, the final model included GSL, nightlights, proximity, and elevation (as mentioned before, proximity was dropped due to high correlation with the elevation covariate). This covariate selection reduced the model AIC from 3,946.6 to 3,911.2.

With respect to Tanzania, in the 2015 and 2010 surveys no geospatial covariate was significant. Vaccination was not measured in the 2008 survey, and therefore the models could not be fit for the vaccination outcome for this survey.

5 DISCUSSION

5.1 Evaluation of Spatial Covariates

A discussion of the specific impacts of the chosen covariates on the prevalence of anemia or receipt of all eight basic vaccinations is beyond the scope of this paper. The spatial covariates chosen for this analysis—EVI, GSL, nightlights, proximity to lake/coast, rainfall, UN population count, and elevation—showed mixed results in their use. Several of these covariates were significantly associated with the anemia outcome, as this is an outcome that we know to be more susceptible to environmental and geophysical forces. Elevation was the covariate most frequently found significant in our analysis (significant in all six surveys), followed by rainfall (four surveys), and EVI, growing season length, and proximity (three surveys each). Except for the Rwanda 2014-15 DHS, in all other surveys the elevation contributed to the largest decrease in the model AIC compared with the AIC of the base model, with rainfall contributing the second-largest decrease, on average. While elevation appeared to be an important environmental variable in predicting the anemia outcome, we believe it is still worth further investigating EVI, GSL, and rainfall, as these covariates are all closely tied to adequate conditions for both the production and availability of agriculture and the propagation and spread of malaria, a known risk factor for the development of anemia (Menon and Yoon 2015; Foote et al. 2013).

These same covariates fared much more poorly when used to analyze receipt of all eight basic vaccinations, a coverage indicator. Only the Burundi 2016-17 DHS and the Uganda 2016 DHS showed any statistically significant spatial covariates (GSL and elevation in Burundi; GSL, nightlights, elevation, and proximity to lake/coast in Uganda). This comes as no real surprise: access to basic vaccinations would be unlikely to be affected by environmental factors, such as rainfall, except in the case of catastrophic extremes—for example, roads entirely cut off by flooding due to heavy rains. However, the significance of the relationship between GSL and, separately, elevation in both Burundi and Uganda to receipt of vaccination, warrants further, country-specific investigation. While length of growing season is associated with prevalence of vector-borne diseases (Tottrup et al. 2009), and rainfall and temperature during the growing season can increase the prevalence of malnutrition (Hagos et al. 2014), growing season could also be associated with immunization coverage. For instance, caregivers could be busy working in fields during longer growing seasons and might not have time to take children to well-child visits where they would get vaccinations. Conversely, shorter growing seasons are associated with increased poverty, resulting in less access to transportation to health services; thus, children's health can be affected.

The results of the temporal analysis of both case studies in Tanzania showed the spatial covariates were better suited for a model with the anemia outcome than the vaccination outcome. Whereas no spatial covariates were significantly associated with receipt of all eight basic vaccinations in the Tanzania 2010 or 2015-16 DHS surveys (vaccination data were not collected as part of the Tanzania 2008 THMIS), most covariates were found significantly associated with anemia in all three surveys. Of note is the Tanzania 2015-16 survey, wherein all the covariates were significant in the analysis of any anemia in children, with elevation and rainfall once again contributing the greatest decreases to the model AIC compared with the AIC of the base model (-47.9 and -46.3, respectively).

5.3 Suggestions for Further Study

This study attempted to demonstrate the feasibility of the DHS Program’s spatial covariate datasets for inclusion in regression modeling, and as such it focused primarily on the model-building process and the quality of the subsequent regression model itself. However, there were several limitations to our approach. Primarily, we were unable to include the malaria covariate in our analysis because it was developed using DHS survey data as well as some of the geospatial variables, such as nighttime lights, which would invariably have led to a circular analysis. This was a disappointment, as ample literature has implicated concomitant malaria infection as a risk factor for development of anemia, and as several covariates included in this study—such as rainfall and EVI, both factors associated with the proliferation of *Anopheles gambiae* mosquitoes—could potentially be serving as corollaries for malaria. Including a spatially linked malaria covariate could potentially help clear up these interactions.

As noted in Section 5.1, the chosen covariates did not model well for a coverage indicator such as receipt of all eight basic vaccinations. Accordingly, the development and use of covariates better suited for coverage indicators is required. The travel times covariate, which measures the amount of time required to reach a settlement of >50,000 people, along with the population density and urbanicity covariates, may have been better fits for modeling vaccination coverage. Further, the creation of covariates measuring the cluster’s proximity to nearest health facility could potentially help define access to health care for individuals, a direct predictor of their access to vaccinations. However, this depends on the existence of master facility datasets—of which we are not currently aware—that can be used for the creation of this dataset. This report was mainly used as a demonstration of the procedures that could be used to include geospatial variables in the analysis of DHS data, and was not intended for finding the best model for the case studies. Future studies on vaccination coverage using geospatial variables may want to consider travel times as another potential covariate to include in their analyses.

Finally, because its scope was limited to an exercise in building a logit regression model, this report does not delve into the specific associations between the covariates and the outcomes. It is our hope that the report demonstrates that using these covariates in such a fashion is a feasible analytical method that can be used by other researchers, who will be able to probe these associations further.

APPENDICES

APPENDIX A DOCUMENTATION FOR INCLUDED COVARIATES

Column Name: Growing Season Length

Derived Data Set: Length of Available Growing Period (16 classes)

Derived Data Set Cell Size: 5 Arc Minute (~0.0833333 decimal degrees; ~10 km)

Derived Data Set License: None (All Rights Reserved)

Year: Based on data collected between 1961 and 1991

Units: Individual classes between 1 and 16. The values are listed below.

1: 0 days	9: 210 - 239 days
2: 1 - 29 days	10: 240 - 269 days
3: 30 - 59 days	11: 270 - 299 days
4: 60 - 89 days	12: 300 - 329 days
5: 90 - 119 days	13: 330 - 364 days
6: 120 - 149 days	14: < 365 days
7: 150 - 179 days	15: 365 days
8: 180 - 209 days	16: > 365 days

Description:

It is impossible to deep link to the dataset. Searching for “growing season” and limiting results to “World” datasets should bring “Length of Available Growing Period (16 classes).”

Length of available growing period refers to the number of days within the period of temperatures above 5° C when moisture conditions are considered adequate. Under rain-fed conditions, the beginning of the growing period is linked to the start of the rainy season. The growing period for most crops continues beyond the rainy season and, to a greater or lesser extent, crops mature on moisture stored in the soil profile.

The mode of the growing season length indices of the cells whose centroid falls within a radius of 10 km (for rural points) or 2 km (for urban points).

Citation:

Food and Agriculture Organization. 2007. “Length of Available Growing Period (16 classes).” Accessed August 21, 2017. <http://www.fao.org/geonetwork/srv/en/main.home>.

Column Name: Enhanced Vegetation Index YEAR

Derived Data Set: Vegetation Index and Phenology (VIP) Phenology EVI-2 Yearly Global 0.05Deg CMG V004

Derived Data Set Cell Size: 0.05 decimal degrees (~5 km)

Derived Data Set License: Public Domain

Year: 1985, 1990, 1995, 2000, 2005, 2010, or 2015

Units: Vegetation index value between 0 (least vegetation) and 10000 (most vegetation)

Description:

The enhanced vegetation index was calculated by measuring the density of green leaves in the near-infrared and visible bands.

The average enhanced vegetation index of the cells whose centroid falls within a radius of 10 km (for rural points) or 2 km (for urban points).

Citation:

Kamel Didan. 2016. "NASA MEaSURES Vegetation Index and Phenology (VIP) Phenology EVI2 Yearly Global 0.05Deg CMG." Accessed August 21, 2017.
https://doi.org/10.5067/measures/vip/vipphen_evi2.004.

Column Name: Proximity to Water

Derived Data Set: GSHHG (Global Self-consistent, Hierarchical, High-resolution Geography Database)

Derived Data Set License: GNU Lesser General Public License

Year: 2017

Units: Meters

Description:

Straight-line distance to the nearest major water body. Based on the World Vector Shorelines, CIA World Data Bank II, and Atlas of the Cryosphere.

Citation:

Wessel, Paul, and Walter Smith. 1996. "A Global Self-consistent, Hierarchical, High-resolution Shoreline Database" *Journal of Geophysical Research* 101:8741-8743. <https://doi.org/10.1029/96JB00104>.

Wessel, Paul, and Walter Smith. 2017. "A Global Self-consistent, Hierarchical, High-resolution Geography Database Version 2.3.7." Accessed August 21, 2017. <http://www.soest.hawaii.edu/pwessel/gshhg/>.

Column Name: Rainfall YEAR

Derived Data Set: Climate Hazards Group InfraRed Precipitation with Station data 2.0

Derived Data Set Cell Size: 0.05 decimal degrees (~5 km)

Derived Data Set License: Public Domain

Year: 1985, 1990, 1995, 2000, 2005, 2010, 2015

Units: Millimeters per year

Description:

The average rainfall of the cells whose centroid falls within a radius of 10 km (for rural points) or 2 km (for urban points).

Citation:

Climate Hazards Group. 2017. "Climate Hazards Group InfraRed Precipitation with Station Data 2.0." Accessed August 21, 2017. <http://chg.geog.ucsb.edu/data/chirps/index.html>.

Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell and Joel Michaelsen. 2015. "The Climate Hazards Infrared Precipitation with Stations—a New Environmental Record for Monitoring Extremes." *Scientific Data* 2. <http://doi.org/10.1038/sdata.2015.66>.

Column Name: Nightlights Composite

Derived Data Set: Version 1 VIIRS Day/Night Band Nighttime Lights

Derived Data Set Cell Size: 15 Arc Second (~0.00416667 decimal degrees; ~500 m)

Derived Data Set License: Public Domain

Year: 2015

Units: Composite cloud-free radiance values

Description:

The average radiance of the cells whose centroid falls within a radius of 10 km (for rural points) or 2 km (for urban points).

Citation:

Mills, Stephen, Stephanie Weiss, and Calvin Liang. 2013. "VIIRS Day/night Band (DNB) Stray Light Characterization and Correction." *Proceedings of SPIE* 8866. <http://dx.doi.org/10.1117/12.2023107>.

National Centers for Environmental Information. 2015. "2015 VIIRS Nighttime Lights Annual Composite." Accessed August 21, 2017. https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html.

Column Name: UN Population Count YEAR

Derived Data Set: UN-Adjusted Population Count, v4 (2000, 2005, 2010, 2015, 2020)

Derived Data Set Cell Size: 30 arc seconds (~0.00833333 decimal degrees; ~1 km)

Derived Data Set License: Non-Standard Noncommercial License

Year: 2000, 2005, 2010, 2015

Units: Number of people

Description:

The average number of people in the cells whose centroid falls within a radius of 10 km (for rural points) or 2 km (for urban points).

Citation:

Center for International Earth Science Information Network – Columbia University. 2016. “Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals.” <http://dx.doi.org/10.7927/H4SF2T42>.

APPENDIX B SUBSET OF RWANDA 2014-15 DHS SPATIAL COVARIATE DATASET

Appendix B Subset of Rwanda 2014-15 DHS spatial covariate dataset

DHSID	GPS_ Dataset	DHS CC	DHS YEAR	DHS cluster	Survey ID	Enhanced_Vegetation_Index_2010	Enhanced_Vegetation_Index_2015	Growing_Season_Length	Nightlights_Composite	Proximity_to_Water	Rainfall_2010	Rainfall_2015	UN_Pop_Count_2010	UN_Pop_Count_2015
RW201400000001	RWGE72FL	RW	2014	1	RW2015DHS	3491.6	3437.9	11	0.0717	35146.9	1268.1	1145.9	170538.5	183040.2
RW201400000002	RWGE72FL	RW	2014	2	RW2015DHS	3641.1	3450.4	11	0.0053	26115.0	1039.6	862.1	144292.3	166278.7
RW201400000003	RWGE72FL	RW	2014	3	RW2015DHS	3797.4	3789.6	13	0.0061	29001.5	1430.1	1232.8	176284.4	187634.7
RW201400000004	RWGE72FL	RW	2014	4	RW2015DHS	3507.2	3416.6	11	0.0004	10328.9	1108.4	895.1	129656.2	184579.6
RW201400000005	RWGE72FL	RW	2014	5	RW2015DHS	3230.8	3003.1	11	0.0263	15599.0	985.3	745.5	82682.2	97288.7
RW201400000006	RWGE72FL	RW	2014	6	RW2015DHS	3053.0	3137.0	11	0.1765	21380.9	1097.0	864.0	9573.9	13104.5
RW201400000007	RWGE72FL	RW	2014	7	RW2015DHS	3855.9	3809.6	13	0.0361	17969.6	1323.6	1084.8	153476.7	153924.8
RW201400000008	RWGE72FL	RW	2014	8	RW2015DHS	3818.3	3656.5	11	0.0357	41680.6	1178.4	990.3	162894.6	182278.9
RW201400000009	RWGE72FL	RW	2014	9	RW2015DHS	3053.0	3137.0	11	5.3833	22203.4	1067.0	830.0	36581.3	41759.3
RW201400000010	RWGE72FL	RW	2014	10	RW2015DHS	3779.0	3682.0	11	1.2922	11274.4	1051.0	856.0	10831.3	12194.1
RW201400000011	RWGE72FL	RW	2014	11	RW2015DHS	3551.1	3407.6	11	0.1689	17244.3	1037.4	842.4	157368.6	192486.2
RW201400000012	RWGE72FL	RW	2014	12	RW2015DHS	3555.8	3527.9	12	0.1722	29085.7	1154.7	998.6	189267.1	212988.7
RW201400000013	RWGE72FL	RW	2014	13	RW2015DHS	3833.7	3852.4	13	0.0020	19472.5	1386.2	1316.6	102849.5	105146.9
RW201400000014	RWGE72FL	RW	2014	14	RW2015DHS	3711.6	3620.3	12	0.0189	25255.9	1307.6	1150.7	125606.9	131618.3
RW201400000015	RWGE72FL	RW	2014	15	RW2015DHS	3577.0	3518.0	13	0.1327	34118.0	1384.0	1201.0	7410.8	8235.6
RW201400000016	RWGE72FL	RW	2014	16	RW2015DHS	2768.7	2693.7	10	0.0017	2015.5	1028.5	833.9	52838.6	61841.3
RW201400000017	RWGE72FL	RW	2014	17	RW2015DHS	3688.6	3677.3	12	0.0108	28709.7	1115.0	919.8	169464.0	194128.3
RW201400000018	RWGE72FL	RW	2014	18	RW2015DHS	3284.0	3315.0	14	0.0956	24186.5	1576.0	1407.0	9856.0	10439.1
RW201400000019	RWGE72FL	RW	2014	19	RW2015DHS	3819.1	3879.3	11	0.0061	8071.9	1124.2	927.5	132309.5	152546.8
RW201400000020	RWGE72FL	RW	2014	20	RW2015DHS	3685.0	3590.9	12	0.0948	34415.9	1213.3	1112.7	147969.5	171832.6
RW201400000021	RWGE72FL	RW	2014	21	RW2015DHS	3724.0	3730.1	13	0.0018	17324.7	1329.9	1224.1	109893.8	114137.9
RW201400000022	RWGE72FL	RW	2014	22	RW2015DHS	3703.5	3710.0	12	0.2033	20059.4	1167.2	1007.0	170443.3	198064.7
RW201400000023	RWGE72FL	RW	2014	23	RW2015DHS	3321.8	3364.8	11	3.1595	25420.4	1107.7	861.8	641618.5	737505.7
RW201400000024	RWGE72FL	RW	2014	24	RW2015DHS	3830.7	3832.9	14	0.1668	10506.8	1389.0	1216.1	378807.9	438056.2
RW201400000025	RWGE72FL	RW	2014	25	RW2015DHS	3245.0	3192.0	14	0.0362	4828.9	1372.0	1131.0	6758.8	6667.9
RW201400000026	RWGE72FL	RW	2014	26	RW2015DHS	3472.3	3370.3	12	0.0026	22541.1	947.2	822.9	36142.3	57533.8
RW201400000027	RWGE72FL	RW	2014	27	RW2015DHS	2451.4	2423.0	14	0.5506	2670.3	1262.6	1120.6	348332.9	411467.2
RW201400000028	RWGE72FL	RW	2014	28	RW2015DHS	3835.2	3806.9	13	0.0155	15307.4	1480.4	1278.9	212410.2	215791.3
RW201400000029	RWGE72FL	RW	2014	29	RW2015DHS	2513.0	2598.0	11	21.8008	30236.8	1084.0	872.0	114594.9	113792.8
RW201400000030	RWGE72FL	RW	2014	30	RW2015DHS	3673.6	3550.2	11	0.0137	28770.6	974.9	793.3	102749.9	128645.6

APPENDIX C ANEMIA MODEL SELECTION – FULL MODEL

Appendix Table C Anemia model selection – Full model for each survey

Variable	Burundi 2016 DHS	Rwanda 2014 DHS	Uganda 2016 DHS	Tanzania 2015 DHS	Tanzania 2010 DHS	Tanzania 2008 DHS
Geospatial variables						
Enhanced Vegetation Index 2015				-0.00002		
Growing Season Length		-0.1157*		0.0546		0.1734***
Nightlights Composite				0.0124	-0.01509	-0.0080
Proximity to Water				0.0000001	0.000000009	0.0000004
Rainfall 2015	-0.0008***		0.0005**	0.0004*		
UN Population Count 2015				-0.00000006		
Cluster Altitude in Meters	-0.0003*	-0.0002	-0.0009***	-0.0003*	-0.00071***	-0.0006***
Enhanced Vegetation Index 2010					-0.00032***	
Enhanced Vegetation Index 2005						0.00002
Rainfall 2005						0.0005
UN Population Count 2005						-0.000003***
Control variables						
<i>Wealth index</i>						
Lowest (Ref.)						
Second	-0.0350	-0.0465	-0.3981**	-0.1528	0.0803	-0.1196
Middle	-0.3360**	-0.1707	-0.6552***	-0.2231*	-0.0974	-0.2426
Fourth	-0.5033***	-0.4301**	-0.6386***	-0.5544***	-0.1669	-0.2135
Highest	-0.7472***	-0.7003***	-0.7845***	-0.7479***	-0.2793	-0.2307
<i>Sex of child</i>						
Male (Ref.)						
Female	-0.1644**	-0.0216	-0.1114	-0.1419**	-0.3008***	-0.1841**
<i>Age of Child</i>						
5-29 months (Ref.)						
30-59 months	-0.4550***	-1.1257***	-1.0797***	-1.1861***	-0.9462***	-1.0317***
<i>Mother's education</i>						
None (Ref.)						
Primary	-0.1838**	-0.1802	-0.4117**	-0.3588***	-0.1085	-0.1860*
Secondary or more	-0.4604***	0.0115	-0.4359*	-0.3957**	-0.1153	0.2732
Observations	5,273	3,144	3,872	7,850	5,906	5,437

*p<0.05, **p<0.01, ***p<0.001

APPENDIX D VACCINATION MODEL SELECTION – FULL MODEL

Appendix Table D Vaccination model selection – Full model for each survey

Variable	Burundi 2016 DHS	Uganda 2016 DHS
Geospatial variables		
Enhanced Vegetation Index 2015		
Growing Season Length	-0.3930***	-0.0917*
Nightlights Composite		-0.0272
Proximity to Water		0.000003
Rainfall 2015		
UN Population Count 2015		
Cluster Altitude in Meters	0.0008***	0.0008***
Enhanced Vegetation Index 2010		
Rainfall 2010		
UN Population Count 2010		
Control variables		
<i>Wealth index</i>		
Lowest (Ref.)		
Second	0.3511*	0.0037
Middle	0.2955	0.0235
Fourth	0.422	-0.0059
Highest	0.3203	0.0906
<i>Sex of child</i>		
Male (Ref.)		
Female	0.1282	-0.0331
<i>Mother's education</i>		
None (Ref.)		
Primary	0.2633*	0.0114
Secondary or more	0.2053	0.3229
Observations	2,585	2,922
*p<0.05, **p<0.01, ***p<0.001		

REFERENCES

- Arab, A., M. C. Jackson, and C. Kongoli. 2014. "Modelling the Effects of Weather and Climate on Malaria Distributions in West Africa." *Malaria Journal* 13:126-126. <https://doi.org/10.1186/1475-2875-13-126>.
- Balarajan, Y., U. Ramakrishnan, E. Ozaltin, A. H. Shankar, and S. V. Subramanian. 2011. "Anaemia in Low-Income and Middle-Income Countries." *Lancet* 378 (9809):2123-35. [https://doi.org/10.1016/S0140-6736\(10\)62304-5](https://doi.org/10.1016/S0140-6736(10)62304-5).
- Burgert-Brucker, C. R., J. Yourkavitch, S. Assaf, and S. Delgado. 2015. *Geographic Variation in Key Indicators of Maternal and Child Health across 27 Countries in Sub-Saharan Africa*. DHS Spatial Analysis Reports No. 12. Rockville, Maryland, USA: ICF International. <http://dhsprogram.com/pubs/pdf/SAR12/SAR12.pdf>.
- Canavan, M. E., H. L. Sipsma, G. M. Kassie, and E. H. Bradley. 2014. "Correlates of Complete Childhood Vaccination in East African Countries." *PLoS ONE* 9(4):e95709. <https://doi.org/10.1371/journal.pone.0095709>.
- Ejigu, B. A., E. Wencheko, and K. Berhane. 2018. "Spatial Pattern and Determinants of Anaemia in Ethiopia." *PLoS ONE* 13(5):e0197171. <https://doi.org/10.1371/journal.pone.0197171>.
- Ernst, K. C., K. A. Lindblade, D. Koech, P. O. Sumba, D. O. Kuwuor, C. C. John, and M. L. Wilson. 2009. "Environmental, Socio-Demographic and Behavioural Determinants of Malaria Risk in the Western Kenyan Highlands: A Case-Control Study." *Tropical Medicine & International Health* 14(10):1258-65. <https://doi.org/10.1111/j.1365-3156.2009.02370.x>.
- Foote, E. M., K. M. Sullivan, L. J. Ruth, J. Oremo, I. Sadumah, T. N. Williams, and P. S. Suchdev. 2013. "Determinants of Anemia among Preschool Children in Rural, Western Kenya." *The American Journal of Tropical Medicine and Hygiene* 88(4):757-764. <https://dx.doi.org/10.4269%2Fajtmh.12-0560>.
- Gari, T., E. Loha, W. Deressa, T. Solomon, H. Atsbeha, M. Assegid, A. Hailu, and B. Lindtjørn. 2017. "Anaemia among Children in a Drought Affected Community in South-Central Ethiopia." *PLoS ONE* 12(3):e0170898. <https://doi.org/10.1371/journal.pone.0170898>.
- Hagos, S., T. Lunde, D. H. Mariam, T. Woldehanna, and B. Lindtjørn. 2014. "Climate Change, Crop Production and Child Under Nutrition in Ethiopia; a Longitudinal Panel Study." *BMC Public Health* 14(1):884. <https://doi.org/10.1186/1471-2458-14-884>.
- Ismael, N., R. Olivier, and K. Barthelemy. 2007. "Risk Factors for Moderate to Severe Anemia among Children in Benin and Mali: Insights from a Multilevel Analysis." *Food and Nutrition Bulletin* 28(1):76-89. <https://doi.org/10.1177/156482650702800109>.
- Janko, M. M., S. R. Irish, B. J. Reich, M. Peterson, S. M. Doctor, M. K. Mwandagalirwa, J. L. Likwela, A. K. Tshefu, S. R. Meshnick, and M. E. Emch. 2018. "The Links between Agriculture, Anopheles Mosquitoes, and Malaria Risk in Children Younger Than 5 Years in the Democratic Republic of the

Congo: A Population-Based, Cross-Sectional, Spatial Study.” *The Lancet. Planetary Health* 2(2):e74-e82. [https://doi.org/10.1016/S2542-5196\(18\)30009-3](https://doi.org/10.1016/S2542-5196(18)30009-3).

Khan, J. R., N. Awan, and F. Misu. 2016. “Determinants of Anemia among 6–59 Months Aged Children in Bangladesh: Evidence from Nationally Representative Data.” *BMC Pediatrics* 16:3. <https://dx.doi.org/10.1186%2Fs12887-015-0536-z>.

Kuziga, F., Y. Adoke, and R. K. Wanyenze. 2017. “Prevalence and Factors Associated with Anaemia among Children Aged 6 to 59 months in Namutumba District, Uganda: A Cross- Sectional Study.” *BMC Pediatrics* 17:25. <https://dx.doi.org/10.1186%2Fs12887-017-0782-3>.

Lakew, Y., A. Bekele, and S. Biadgilign. 2015. “Factors Influencing Full Immunization Coverage among 12–23 Months of Age Children in Ethiopia: Evidence from the National Demographic and Health Survey in 2011.” *BMC Public Health* 15(1):728. <https://doi.org/10.1186/s12889-015-2078-6>.

Legason, I. D., A. Atiku, R. Ssenyonga, P. Olupot-Olupot, and J. B. Barugahare. 2017. “Prevalence of Anaemia and Associated Risk Factors among Children in North-Western Uganda: A Cross Sectional Study.” *BMC Hematology* 17:10. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5496358>.

Liu, L., S. Oza, D. Hogan, J. Perin, I. Rudan, J. E. Lawn, S. Cousens, C. Mathers, and R. E. Black. 2015. “Global, Regional, and National Causes of Child Mortality in 2000-13, with Projections to Inform Post-2015 Priorities: An Updated Systematic Analysis.” *Lancet* 385(9966):430-40. [https://doi.org/10.1016/S0140-6736\(14\)61698-6](https://doi.org/10.1016/S0140-6736(14)61698-6).

Mboera, L. E., K. P. Senkoro, B. K. Mayala, S. F. Rumisha, R. T. Rwegoshora, M. R. Mlozi, and E. H. Shayo. 2010. “Spatio-Temporal Variation in Malaria Transmission Intensity in Five Agro-Ecosystems in Mvomero District, Tanzania.” *Geospatial Health* 4(2):167-78. <https://doi.org/10.4081/gh.2010.198>.

Menon, M. P., and S. S. Yoon. 2015. “Prevalence and Factors Associated with Anemia among Children under 5 Years of Age—Uganda, 2009.” *The American Journal of Tropical Medicine and Hygiene* 93(3):521-526. <https://dx.doi.org/10.4269%2Fajtmh.15-0102>.

Ncogo, P., M. Romay-Barja, A. Benito, P. Aparicio, G. Nseng, P. Berzosa, M. A. Santana-Morales, M. Riloha, B. Valladares, and Z. Herrador. 2017. “Prevalence of Anemia and Associated Factors in Children Living in Urban and Rural Settings from Bata District, Equatorial Guinea, 2013.” *PLoS ONE* 12(5):e0176613. <https://doi.org/10.1371/journal.pone.0176613>.

Ngesa, O., and H. Mwambi. 2014. “Prevalence and Risk Factors of Anaemia among Children Aged between 6 Months and 14 Years in Kenya.” *PLoS ONE* 9(11):e113756. <https://doi.org/10.1371/journal.pone.0113756>.

Ricotta, E. E., S. A. Frese, C. Choobwe, T. A. Louis, and C. J. Shiff. 2014. “Evaluating Local Vegetation Cover as a Risk Factor for Malaria Transmission: A New Analytical Approach Using ImageJ.” *Malaria Journal* 13(1):94. <https://doi.org/10.1186/1475-2875-13-94>.

Sewe, M., J. Rocklöv, J. Williamson, M. Hamel, A. Nyaguara, F. Odhiambo, and K. Laserson. 2015. “The Association of Weather Variability and Under Five Malaria Mortality in KEMRI/CDC HDSS in

- Western Kenya 2003 to 2008: A Time Series Analysis.” *International Journal of Environmental Research and Public Health* 12(2):1983-1997. <https://doi.org/10.3390/ijerph120201983>.
- Simbauranga, R. H., E. Kamugisha, A. Hokororo, B. R. Kidenya, and J. Makani. 2015. “Prevalence and Factors Associated with Severe Anaemia Amongst Under-Five Children Hospitalized at Bugando Medical Centre, Mwanza, Tanzania.” *BMC Hematology* 15:13. <https://dx.doi.org/10.1186%2Fs12878-015-0033-5>.
- Tottrup, C., B. P. Tersbol, W. Lindeboom, and D. Meyrowitsch. 2009. “Putting Child Mortality on a Map: Towards an Understanding of Inequity in Health.” *Tropical Medicine & International Health* 14(6):653-62. <https://doi.org/10.1111/j.1365-3156.2009.02275.x>.
- Walsh, J. F., D. H. Molyneux, and M. H. Birley. 1993. “Deforestation: Effects on Vector-Borne Disease.” *Parasitology* 106 Suppl:S55-75.
- Warren, J. L., C. Perez-Heydrich, C. R. Burgert, and M. E. Emch. 2016. “Influence of Demographic and Health Survey Point Displacements on Distance-Based Analyses.” *Spatial Demography* 4(2):155-173. <https://doi.org/10.1007/s40980-015-0014-0>.
- WHO. 2017a. *1 in 10 Infants Worldwide Did Not Receive Any Vaccinations in 2016*. World Health Organization. <http://www.who.int/news-room/detail/17-07-2017-1-in-10-infants-worldwide-did-not-receive-any-vaccinations-in-2016>.
- WHO. 2017b. *Nutritional Anaemias: Tools for Effective Prevention and Control*. World Health Organization. <http://www.who.int/nutrition/publications/micronutrients/anaemias-tools-prevention-control/en/>.
- WHO. 2018. *Immunization Coverage Fact Sheet*. World Health Organization. <http://www.who.int/en/news-room/fact-sheets/detail/immunization-coverage>.