



Contextual data in IPUMS DHS: physical and social environment variables linked to the Demographic and Health Surveys

Elizabeth Heger Boyle¹ · Miriam L. King² · Sarah Garcia¹  · Corey Culver¹ · Jordan Bourdeaux¹

Published online: 27 May 2020
© Springer Nature B.V. 2020

Abstract

The Demographic and Health Surveys (DHS) are the most important source of comparative information on the health of women and young children in low- and middle-income countries and are well-suited for studies of the relationship between environmental factors and health. However, barriers have limited the use of the DHS for these purposes. IPUMS DHS, an online data dissemination tool, overcomes these barriers, simplifying comparative analyses with DHS. IPUMS DHS recently incorporated environmental variables that can easily be attached to individual or household records, facilitating the use of DHS data for the study of population and environment issues. We provide a brief introduction to IPUMS DHS, describe the current and anticipated environmental variables and how to use them, and provide an example of

✉ Sarah Garcia
sgarcia@umn.edu

Elizabeth Heger Boyle
boyle014@umn.edu

Miriam L. King
kingx025@umn.edu

Corey Culver
culve051@umn.edu

Jordan Bourdeaux
boudr053@umn.edu

¹ Sociology Department, 909 Social Sciences Building, University of Minnesota, 267 19th Avenue South, Minneapolis, MN 55455, USA

² Institute for Social Research and Data Innovation, University of Minnesota, Minneapolis, MN 55455, USA

the novel research possibilities facilitated by this latest IPUMS DHS development. IPUMS-DHS is available free online at dhs.ipums.org.

Introduction

The Demographic and Health Surveys (DHS) are the most important source of comparative information on the health of non-elderly adults and young children in low- and middle-income countries (LMICs) (Corsi, Neuman, Finlay, & Subramanian, *International Journal of Epidemiology*, 41, 1602–1613, Corsi et al. 2012). DHS data (<https://dhsprogram.com/publications/Journal-Articles-Search.cfm>) are the basis for over 6000 journal articles since 1986. Since 1984, with funding from the US Agency for International Development (USAID), the DHS has asked women of childbearing age in LMICs about their personal backgrounds, health, birth histories, and the health of their young children. The surveys also collect data on household characteristics and household members. In the 1990s, they began collecting information from men. A call for greater attention to context (Butz and Torrey 2006; Matthews 2008) prompted The DHS Program to provide GPS location information for sampling clusters, somewhat displaced to maintain confidentiality, for most of their surveys.

In 2012, a team at the University of Minnesota, with funding from NICHD, created IPUMS DHS (Boyle et al. 2019) to facilitate comparative over-time and cross-national research with the DHS. IPUMS DHS harmonizes variable names and codes across surveys, makes it unnecessary to merge files, and allows researchers to download a single integrated data file, customized to include only selected variables. In IPUMS DHS, each variable name links to extensive documentation, such as the related survey question and questionnaire skip pattern. IPUMS DHS offers free access to thousands of harmonized DHS variables across many samples.

Exploiting georeferenced information in the DHS's public use GPS datasets, IPUMS DHS now provides contextual variables linked to individual records based on sample cluster locations. These variables cover physical environment, economic and population features, and agriculture. With this latest capability, researchers without geographic training can easily use contextual variables with DHS data. Contextual variables constructed around DHS's cluster location points reflect subjects' relatively close physical environment and better accommodate heterogeneity in environmental variables than some alternative sources, such as IPUMS Terra, which supplies combined land use, climate, and population data for larger geographic areas.

IPUMS DHS currently includes the full range of DHS variables for 38 countries and 156 samples, with most standard DHS samples from Africa. IPUMS DHS has also created contextual variables for **all** DHS files with GPS location data, shown in Table 1. The IPUMS DHS website offers the contextual variables in downloadable .csv files that can be merged with DHS public use files. Samples **bolded** in Table 1 are fully integrated into IPUMS DHS. For these bolded samples, researchers can create a customized data file of harmonized variables, including any relevant contextual variables, without having to merge files. We describe both processes below.

IPUMS DHS incorporates new material every year, adding DHS surveys from more countries and the latest surveys from countries already in the database. (Our funding prioritizes harmonizing standard DHS samples from Africa, the Middle East, and South Asia.) For any newly released DHS GPS files, we create new .csv files of contextual

Table 1 DHS Samples with GPS Data, as of August 2019

Country	DHS sample years
Albania	2008
Angola	2006, 2011, 2015
Armenia	2010, 2015
Bangladesh	2000, 2004, 2007, 2011, 2014
Benin	1996, 2001, 2012, 2017
Bolivia	2008, 2010
Burkina Faso	1993, 1999, 2003, 2010, 2014
Burundi	2010, 2012, 2016
Cambodia	2000, 2005, 2010, 2014
Cameroon	1991, 2004, 2011
Central African Republic	1994
Chad	2014
Colombia	2010
Comoros	2012
Congo, Democratic Republic	2007, 2013
Cote D'Ivoire	1994, 1998, 2012
Dominican Republic	2007, 2013
Egypt	1992, 1995, 2000, 2003, 2005, 2008, 2014
Ethiopia	2000, 2005, 2010, 2016
Gabon	2012
Ghana	1993, 1998, 2003, 2008, 2014, 2016
Guatemala	2015
Guinea	1999, 2005, 2012
Guyana	2009
Haiti	2000, 2006, 2012, 2017
Honduras	2011
India	2015
Indonesia	2003
Jordan	2002, 2007, 2012, 2017
Kenya	2003, 2008, 2014, 2015
Kyrgyzstan	2012
Lesotho	2004, 2009, 2014
Liberia	1986, 2007, 2009, 2011, 2013, 2016
Madagascar	1997, 2008, 2011, 2013, 2016
Malawi	2000, 2004, 2010, 2012, 2014, 2015, 2016, 2017
Mali	1996, 2001, 2006, 2010, 2012, 2015
Moldova	2005
Morocco	2003
Mozambique	2009, 2011, 2015, 2018
Myanmar	2015
Namibia	2000, 2006, 2013
Nepal	2001, 2006, 2011, 2016
Niger	1992, 1998

Table 1 (continued)

Country	DHS sample years
Nigeria	1990, 2003, 2008, 2010, 2013, 2015
Pakistan	2006
Peru	2000, 2004, 2009
Philippines	2003, 2008, 2017
Rwanda	2005, 2008, 2010, 2014
Senegal	1993, 1997, 2005, 2008, 2010, 2012, 2015, 2016
Sierra Leone	2008, 2013, 2016
Swaziland	2006
Tajikistan	2012
Tanzania	1999, 2003, 2007, 2010, 2012, 2015, 2017
Timor-Leste	2009, 2016
Togo	1988, 1998, 2013
Uzbekistan	2002
Zambia	2007, 2013
Zimbabwe	1999, 2005, 2010, 2015

Researchers can download .csv files of contextual variables for all these samples from the IPUMS DHS website. **Bolded** samples also have integrated DHS variables available through IPUMS DHS

variables. Ultimately, we expect to harmonize the full range of DHS data, including AIS (AIDS monitoring) and MIS (malaria monitoring) surveys. We also plan to add other contextual variables, and we invite suggestions for including other high-quality measures available for a broad range of LMICS.

IPUMS DHS contextual variables unlock exciting research opportunities on population and environment, but they are not the best tool for every project. The IPUMS DHS team made choices, described below, about how to deal with the displacement of the public use DHS cluster location information; some scholars might prefer a different solution. Additionally, we chose which environmental variables to include and, when relevant, how to aggregate them. Researchers who prefer other variables, or who need the variables aggregated differently, may prefer to do their own linking across the data sources. Below, we describe our choices so analysts can make an informed assessment of whether IPUMS DHS is the best fit for their research.

In the following sections, we first describe our methods for dealing with data displacement and constructing IPUMS DHS contextual variables. We next offer information about each contextual variable, with examples of research joining contextual information with DHS data and other sources. We then give guidance on using the IPUMS DHS contextual variables. After providing an example using some of the contextual variables, we describe the limitations of IPUMS DHS and our plans for adding more contextual variables to IPUMS DHS.

Methodology

Displacement The DHS Program displaces public use GPS data points to preserve confidentiality (Burgert et al. 2013). Georeferenced DHS data aggregate cases from

each sampling cluster to a single point coordinate and then mask the coordinate via displacement. Displacement is 0 to 2 km for urban areas and 0 to 5 km for rural areas, with uniformly distributed displaced distances. One percent of rural clusters are displaced up to 10 km. The DHS Program recommends averaging supplemental environmental data over a 5- to 10-km buffer around rural cluster points, though the scale of spatial processes and similarity of nearby cells may influence buffer zone size (Perez-Heydrich et al. 2016).

To deal with displacement, we constructed 5- or 10-km buffer zones around GPS sample points to create IPUMS DHS contextual variables. The creation of most IPUMS DHS contextual variables is based on a common general methodology using Esri's ArcGIS software suite.¹ We constructed 5-km buffer zones around sample points when working with environmental data with clear map boundaries (i.e., population density, soil type, livelihood zone, and ecoregion), with the predominant value in the buffer assigned to each sampled person within the sample cluster. For soil type, livelihood zone, and ecoregion, categorical values are homogeneous within boundaries (e.g., a single livelihood type is assigned to a clearly mapped area within a country). This homogeneity reduces the likelihood of assigning an erroneous value due to missing the sample point for the few rural clusters displaced more than 5 km.

We constructed 10-km buffer zones around sample points for the other contextual variables because the environmental source data had less clearly demarcated boundaries (e.g., pixelated raster data from satellite imageries) and had a range of measured values within those rough boundaries (e.g., for temperature, precipitation, and malarial incidence). For these contextual variables, values—the mean, maximum, sum, or average—were computed for all pixels in a 10-km buffer, and these estimates were assigned to all individuals in each cluster location. A different methodology used for the conflict variables (battles, riots, and violence against civilians) is described below.

Using 5- or 10-km buffer zones around reported GPS sample points matches both the general guidelines for computing environmental variables put forward by The DHS Program and the practice of several publications pairing DHS data with supplemental environmental data.² There are, however, trade-offs in opting for smaller versus larger buffer zones. A larger buffer zone (e.g., 10-km radius) may conceal underlying heterogeneity within small areas; a smaller buffer zone (e.g., 5-km radius) may exclude the actual cluster location for the 1% of rural clusters displaced up to 10 km. Some researchers studying rare events, such as political conflicts, have used larger buffers to include enough cases for analysis.³ Nor is the approach of constructing a buffer zone of at least 5 km around a reported cluster location the only possible strategy. Grace and

¹ Much of the source data was acquired as raster or vector files. Source raster files with a resolution greater than 500 m and vector files were converted to raster files with a resolution of 500 m. We used the Focal Statistics tool to update each pixel value within a 5- or 10-km circular buffer around the sample cluster GPS location.

² For studies pairing NDVI data with DHS data, two publications (Grace et al. 2017; Davenport et al. 2017) used a 10-km buffer zone, while two others (Ickowitz et al. 2014; Brown et al. 2014) used a 5-km buffer zone. For both the 5- and 10-km buffer zones, these authors cited DHS recommendations and typical practice to justify their choice. A study pairing precipitation data with DHS data (Davenport et al. 2017) used a 5-km buffer zone. A study pairing temperature data with DHS data (Grace et al. 2012) used 10-km buffer zones. That some authors use 5- versus 10-km buffer zones in different articles using the same environmental sources suggests an openness to using either of these scales with DHS data.

³ For example, Pelillo (2012) used ACLED data with DHS data using buffer zones of 50, 100, and 200 km.

co-authors recently proposed “that the user selects a settlement near the DHS’ published location and measures the environmental conditions using a buffer much smaller than 10 km ... [hypothesizing that] a small, precise buffer around an incorrect settlement is a better measure of truth than is an overly large buffer around the published point” (Grace et al. 2019). Scholars must determine if the IPUMS DHS approach is suitable for their research.

Selection of contextual variable datasets When selecting which environmental datasets (described below) to pair with the DHS, we generally chose those that were most up-to-date and had the highest spatial and temporal resolution. Among similar variables, we also chose those that were measured directly over those that were imputed.

Variables of physical and environmental setting

IPUMS DHS currently includes contextual variables on the physical environment, economic and population features, and agriculture. Table 2 provides summary information about these contextual variables, including each variable’s meaning, name, original source, temporal resolution, spatial resolution in kilometers in the original data, and variable type (categorical, bounded numeric [from 0 to 1], or unbounded numeric). Every year, IPUMS DHS adds contextual variables for any newly released DHS public use data files with GPS data.⁴

Two environmental variables, **ecoregion** and **soil**,⁵ have categorical values from single points in time and homogeneous values within clearly defined boundaries. Ecoregion is the predominant terrestrial ecoregion (circa 2001) within a 5-km circular buffer around DHS cluster locations. This variable uses a 5-digit coding system consisting of the 1-digit real (ecozone) code, the 2-digit code for biome (habitat type), and the 2-digit specific ecoregion number. There are over 800 terrestrial ecoregions, and users can conduct analyses at any of the three nested levels. Obtained from the Terrestrial Ecoregions of the World database by the World Wildlife Foundation (Olson et al. 2001), ecoregion data are available for all countries. With a few exceptions (e.g., León, 2012), researchers employing DHS data have made little use of data on ecoregions, leaving open many research opportunities.

The variable **soil** specifies the predominant soil type and class for a 5-km circular buffer around DHS clusters. Data covering 118 soil types and 20 top-level categories come from the World Soil Information database (Hengl et al. 2017). Soil data are available for all countries, based on information published circa 2016. Studies have used soil type to explain variation in hookworm parasite

⁴ Once IPUMS DHS releases contextual variables for a given sample, the values for contextual variables for that sample are not recalculated, even if additional historical information is added by the producers of the original supplemental dataset (e.g., for ACLED data on conflicts). Revising the contextual variables after their original release would undermine the scientific replicability of publications incorporating contextual variables from IPUMS DHS. When new supplemental data become public (e.g., observed rather than estimated population counts for the year 2020), IPUMS DHS will calculate and release new contextual variables based on more recent data.

⁵ IPUMS DHS variable names are listed in bold font throughout the article.

Table 2 Summary of contextual variables in IPUMS DHS

Source	Variable	Description	Temporal resolution	Spatial resolution (~km) of source area	Variable type	Buffer (km)	Date Range
<i>Physical and Environmental Context</i>							
Terrestrial Ecoregion of the World (TEOW) SoilGrids, World Soil Information (ISRIC)	ecoregion	Predominant terrestrial ecoregion	NA	NA	Categorical	5	2001
	soil	Soil type	NA	0.25	Categorical	5	2016
	ndvi	Maximum value: Greenness of vegetation and land cover in area	72 months	0.25	Numeric (0-1)	10	February 2000 - October 2018
Climate Hazards Group (CHG) from InfraRed Precipitation with Station data (CHIRPS)	precip	Average monthly precipitation	72 months	5	Numeric	10	January 1981 – October 2018
Terrestrial Hydrology Research Group, Princeton University (version 3)	temp_min temp_max	Minimum and maximum monthly temperature (kelvin scale)	72 months	25	Numeric	10	January 1980 – October 2018
<i>Economic and Social Context</i>							
Famine Early Warning System Network (FEWS NET)	livelihood¹	Predominant livelihood zone (i.e., main economic basis for subsistence)	NA	NA	Categorical	5	2012
Gridded Population of the World (GPW), v4, from Socioeconomic Data and Applications Center (SEDAC)	popdensity	Population density (persons per square kilometer)	Estimates for 2000, 2005, 2010, 2015, 2020	1	Numeric	5	2000 - 2020
The Malaria Atlas Project (MAP)	malaria²	Mean clinical Plasmodium falciparum malaria incidence (expressed as fractions of cases per person)	Yearly data from 2000 - 2015	5	Numeric	10	2000 – 2015
Armed Conflict Location and Event Data Project (ACLED)	battles³	Number of days with battles (yearly)	Varies by country. Longest series, 1997 to present. See IPUMS DHS for details.	NA	Numeric	10	Varies
ACLED	riots³		Varies by country.	NA	Numeric	10	Varies

Table 2 (continued)

Source	Variable	Description	Temporal resolution	Spatial resolution (~km) of source area	Variable type	Buffer (km)	Date Range
ACLED	civ_violence³	Number of days with riots or protests (yearly) Number of days with violence against civilians (yearly)	Varies by country.	NA	Numeric	10	Varies
<i>Agricultural Context</i>							
EarthStat	cropland	Proportion of land used for planting crops	Circa 2000	10	Numeric (0-1)	10	2000
EarthStat	pastureland	Proportion of land used as pasture	Circa 2000	10	Numeric (0-1)	10	2000
EarthStat	Crops: Harvested Area (e.g., rice_h)	Total harvested area in hectares on a subset of 17 major crops. ⁴	2000	10	Numeric	10	2000
EarthStat	Crops: Production (e.g., rice_p)	Total production of a specific crop, in metric tons on a subset of 17 major crops.	2000	10	Numeric	10	2000

¹ LIVELIHOOD ZONE data are country specific (not comparable across countries) and are available for the following countries: Afghanistan, Angola, Burkina Faso, Burundi, Central African Republic, Chad, Democratic Republic of the Congo, Djibouti, El Salvador, Ethiopia, Guatemala, Guinea, Haiti, Honduras, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mozambique, Nicaragua, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, Somalia, South Sudan, Tanzania, Tajikistan, Uganda, Yemen, Zambia, Zimbabwe

² MALARIA data available for the following countries: Angola, Benin, Burkina Faso, Burundi, Central African Republic, Chad, Comoros, Cote d'Ivoire, Democratic Republic of the Congo, Ethiopia, Gabon, Ghana, Guinea, Kenya, Liberia, Madagascar, Malawi, Mozambique, Namibia, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, Tanzania, Togo, Zambia, Zimbabwe

³ For Battles, Riots and Violence, "present" refers to the date for last data download, currently, September 15, 2018.

⁴ The 17 major crops include: Barley, cassava, cotton, groundnut, maize, millet, oil palm, potato, rapeseed, rice, sorghum, soybean, sugar beet, sugarcane, sunflower, wheat, and rye.

prevalence (Mabaso 1998; Mabaso et al. 2003), estimate population-level micro-nutrient intake and deficiency risks (Joy et al. 2015), link agroecosystems to child stunting (Alemu et al. 2017), explain district differences in Indian women's agricultural work and child sex ratios (Carranza 2014), and, together with DHS data, link agricultural capacity to child hunger (Balk et al. 2005).

Three sets of additional environmental variables (greenness, precipitation, and temperature) vary over time and are presented for a total of 72 months around a given DHS survey. By offering 72 months of data, IPUMS DHS has included information most likely relevant to researchers. DHS surveys generally collect information on the health and survival of children born in the 60 months before the interview date,⁶ and some surveys are fielded over several months. The 72 months include measurements for each of 60 months prior to the survey start date, the month of the survey start, and each of 11 months following the survey start date.

NDVI (Normalized Difference Vegetation Index) is a remotely sensed indicator of live green vegetation, with values potentially ranging from negative one to positive one. Because IPUMS DHS provides the maximum observed value across all the pixels within a 10-km circular buffer around each DHS cluster, **the NDVI values in IPUMS DHS are never less than zero**. Barren areas of snow, sand, or rock have the lowest NDVI values; shrub and grassland have moderate NDVI values; and temperate and tropical rainforests have the highest NDVI values. These data can identify areas under stress (such as extended drought) or undergoing change (such as deforestation) (Weir and Herring 2013). Obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) (Didan 2015) and covering all countries, the original data are based on satellite images collected every 16 days, from February 2000 to 2018. The contextual variables on NDVI in IPUMS DHS include 72 variables covering each of 60 months prior to the survey start, the first survey month, and each of the following 11 months.

Multiple publications have used NDVI data with DHS data. Studies have focused on interactions between water and food security for child health (Grace et al. 2017), relationships between agriculture and fertility (Grace and Nagle 2015), mosquito habitat suitability and mosquito net distribution (Acheson 2015), dietary quality and tree cover (Ickowitz et al. 2014), climatic conditions during gestation and infancy and sex-specific vulnerability to undernutrition (Mulmi et al. 2016), effects of environmental change on health (Brown et al. 2014; Johnson and Brown 2014), and the association between climate variability, birthweights, and agricultural livelihoods (Bakhtsiyarava et al. 2018).

Precip (monthly precipitation in milliliters) is reported for a 72-month time series for a 10-km circular buffer around each DHS cluster. Precipitation data were originally obtained from Climate Hazards Group, which provides infraRed Precipitation with Station data (CHIRPS) (Funk et al. 2015). Spanning all longitudes, CHIRPS data constitute a 30+ year quasi-global rainfall dataset that incorporates 0.05° resolution satellite imagery with in situ station data, to create gridded rainfall time series. The **precip** variables are available for all DHS samples with publicly released cluster data (as shown

⁶ Data on the 60 months prior to a survey provide information on climatic conditions during the lives of children under age 5 in that survey but admittedly do not capture conditions when the earliest-born young children were in utero. If we learn that many users desire this further information, we will reconsider this time frame and potentially add information for an additional 9 months before the survey start date.

in Table 1). Researchers pairing precipitation and DHS data have studied the effect of climate change on child health (Davenport et al. 2017; Grace et al. 2015; MacVicar 2016) and the effect of early-life drought exposure on later child stunting (Lillepold et al. 2018) and wealth in adulthood (Hyland and Russ 2019). IPUMS DHS currently includes raw precipitation data for individual months; we plan to add summary monthly precipitation data averaged across decades for comparison purposes in the near future.

Temp_min and **temp_max** include 72 months of temperature data in the 10-km radius of sample clusters. Values are available for all DHS samples with GPS cluster data, based on data for January 1980 to September 2018 from the Terrestrial Hydrology Research Group at Princeton University (Sheffield et al. (2006). Studies focusing on the interplay between climate and health often use temperature data along with precipitation data. Research has found prolonged exposure to high temperatures to be an independent health risk (Grace et al. 2015). IPUMS DHS plans to supplement the existing temperature data for individual months by adding, in the near future, summary monthly temperature data averaged across decades for comparison purposes.

Variables of economic and social contexts

Livelihood reports the predominant economic basis for subsistence within a 5-km circular buffer around DHS clusters. This variable, which has a homogeneous value assigned to a clearly defined subregion within a country, indicates sources of food and/or income and people's vulnerability to hazards such as floods and droughts. Values consist of a 3-digit country code followed by a 3-digit numeric livelihood (e.g., river fishing and staple crops) code. The non-time-variant source data (published in 2012) were obtained from the Famine Early Warning System Network and provide information on 36 countries with DHS samples with cluster location data (Livelihoods 2020).

The livelihood variable is country specific.

Data on livelihood zones combined with DHS data have yielded insights into population vulnerability and human health in Africa. Jankowska et al. (2012) examined current links between livelihood and measures of malnutrition and projected future outcomes (malnutrition and anemia cases), given climate change. Grace and her co-authors incorporated livelihood zone data into multilevel models of staple food prices and birth weights (Grace et al. 2014) and of climate variables and child stunting (Grace et al. 2012) and examined how weather effects on birth weight vary across livelihood zones (Grace et al. 2015). Policy-relevant DHS-based studies have considered how livelihood zones should be factored into efforts to reduce child stunting (Woodruff et al. 2017) and into development efforts (Malcomb et al. 2014).

Popdensity reports persons per square kilometer within a 5-km circular buffer around DHS clusters for the years 2000, 2005, 2010, 2015, and 2020. The data were obtained from the Gridded Population of the World (GPW-v4), provided by the Socioeconomic Data and Applications Center (Center for International Earth Science Information Network (CIESIN) Columbia University, 2016) and are based on counts consistent with national censuses and population registers. A proportional allocation gridding algorithm is used to assign population values to 1-km grid cells, and the population density grids are created by dividing the population count grids by the land

grids.⁷ The original DHS public use files include a widely used dichotomous urban/rural variable, with the definition of “urban” set by each country’s national statistical office.⁸ The **popdensity** variable provides a finer level of resolution that allows researchers to incorporate their own definitions of urban and rural depending on research needs. We use this variable in a data analysis example below.

Malaria consists of 16 separate variables, one value for each year from 2000 to 2015. **Malaria** reports the mean clinical *Plasmodium falciparum* malaria incidence within a 10-km circular buffer around DHS clusters. The malaria incidence is expressed as fractions of cases per person; values range from zero to one. Data were obtained from the Malaria Atlas Project (MAP) (Bhatt et al. 2015) and were expressed as malaria incidence raster layers, with circa 5-km spatial resolution in the World Geodetic System-1984 Geographic Datum coordinate system (National Geospatial-Intelligence Agency 2012). The source data cover 48 countries in sub-Saharan Africa, 30 of which have DHS samples with GPS cluster data.

Published studies have used mapped data on malaria incidence in conjunction with DHS data. Amoah et al. (2018) coupled data on child undernutrition from 20 DHS surveys with MAP data on malaria incidence. Cuadros and co-authors (Cuadros et al. 2011a, 2011b) studied the relationship between malaria and HIV prevalence in East and West Africa, using DHS and MAP data. Kazembe et al. (2007) conducted a spatial analysis of the relationship between early childhood mortality in the DHS and malaria endemicity. More commonly, researchers use data from The DHS Program’s Malaria Indicator Surveys (MIS), which include blood draws for testing for malaria parasites. MAP data on malaria incidence cover additional years and some countries without MIS surveys, however.

Conflict variables (**civ_violence**, **riots**, and **battles**) draw data from the Armed Conflict Location and Event Data Project (ACLED) (Raleigh et al. 2010).⁹ Conflict variables represent an annual count of days with armed or civilian conflicts, according to three conflict types (violence against civilians, riots, and battles), occurring within a 10-km circular buffer around a DHS cluster location. The longest time series cover 1997 to 2016, with longer spans for Africa than for Asia and the Middle East. IPUMS DHS staff converted longitude and latitude coordinates to a point layer, created a 10-km buffer around each DHS sample cluster location, and counted conflict events falling within the buffer in a year. If a buffer crossed an international boundary, only events occurring in the same country as the DHS sample were included. Conflict variables are available for 42 currently available DHS samples.

Studies combining DHS and ACLED data have examined the effect of armed conflict on child welfare in Cote d’Ivoire (Wayoro 2018), the relationship between access to healthcare and civil conflict within African states (Rohde 2014),

⁷ We chose to use Gridded Population of the World data, rather than WorldPop data, because GPW-v4 does not rely on ancillary sources, such as land use and transportation data, when allocating population data from census units to grid cells. If researchers choose to use other data on transportation or land use in conjunction with GPW-4 data, they do not face endogeneity problems. Some researchers understandably consider WorldPop results more reliable due to its reliance on ancillary data and may prefer to construct their own measures of population density using WorldPop.

⁸ According to records of how frequently variables are downloaded, the DHS’s original **urban** (V025) variable on urban versus rural residence is among the most widely used variables. **Popdensity** supplies a supplemental measure that is comparable across countries.

⁹ Another source of conflict data is the Uppsala Conflict Data Program (UCDP; Cunningham et al. 2009), which measures fatalities in countries at war. We plan to incorporate the UCDP into IPUMS DHS in the near future.

negative effects of exposure to violent and non-violent civil insecurity on child growth in rural Africa (Darrouzet-Nardi 2016), and the relationship between past exposure to violent conflict and risky sexual behavior in Uganda (Delavande and Menezes Cordeiro 2012). We provide a further example of this fruitful combination of data from DHS and ACLED from our own research later in this paper.

Agricultural context variables

Cropland and **pastureland** variables provide data as bounded numeric values (from 0 to 1) representing the proportion of a 10-km circular buffer around each DHS cluster that is covered by cropland and pastureland, respectively. The original cropland and pasture area raster layers were acquired from EarthStat (Monfreda et al. 2008; Ramankutty et al. 2008) and cover all countries, based on material from circa 2000 from agricultural inventories and satellite-derived land cover data.

Crop harvest area and crop production variables report area (hectares) dedicated to growing and yield (metric tons) for specific crops, within a 10-km circular buffer around DHS clusters. Data were obtained from EarthStat, which produces data for 175 crops (Monfreda et al. 2008). IPUMS DHS uses a subset of 17 major crops (i.e., barley, cassava, cotton, groundnuts, maize, millet, oil palm, potatoes, rapeseed, rice, rye, sorghum, soybeans, sugar beets, sugarcane, sunflowers, and wheat). EarthStat describes these data as “created by combining national, state, and county level census statistics with a recently updated global data set of croplands” on a 10×10 kilometer latitude/longitude grid, covering all countries circa 2000.

Publications have linked anthropometric measures of undernutrition with indicators of agricultural productivity (Slavchevska 2015; Headey and Hoddinott 2016; Liu et al. 2008; Vepa et al. 2016), examined the relationship between diversity in crops grown and diversity in children’s diets (Hirvonen and Hoddinott 2017; Jones et al. 2014; Rosenberg et al. 2018), and related children’s school attendance and work to the seasonal demands of staple crops (Mohammed 2016). The agricultural data sources in these works are generally idiosyncratic and country specific. The contextual agricultural variables in IPUMS DHS are standard across multiple countries, with fine geographic detail, and are easily combined with DHS data on undernutrition, dietary intake, child labor, and school attendance.

Accessing and using the IPUMS DHS contextual variables

IPUMS DHS data are free to all registered DHS users. Even without registering, anyone can browse the samples, variables, and extensive documentation at the IPUMS DHS website. For DHS samples that are fully integrated and included in the IPUMS DHS database (see bolded entries in Table 1), users select contextual variables for inclusion in their customized data files. Researchers unfamiliar with the IPUMS data dissemination system may want to review the IPUMS DHS “User Guide” or YouTube videos, which explain the intuitive IPUMS interface (click “HELP” at the top of the page at dhs.ipums.org).

Adding a variable to one's data file (referred to as the "Data Cart") adds a single variable to each record for contextual variables with values measured in 1 year and invariant across DHS sample years, such as **soil**. For contextual variables measured at multiple time points, such as **battles** or **popdensity**, adding the variable to one's data file adds the entire time series of variables to each record. For example, a researcher selecting **battles** will receive a series of 20 variables (**battles_1997**, **battles_1998**, and so forth up to **battles_2016**). She must construct a variable for the values useful to her analysis (e.g., **battles** in the year prior to a survey).

To use contextual variables for DHS samples not currently integrated into IPUMS DHS, researchers follow a different process. They can download the .csv files of the contextual variables at dhs.ipums.org (click on "CONTEXTUAL VARIABLES" in the left sidebar), for all of the 168 DHS samples with GPS location data. Researchers then link .csv file data to original DHS files downloaded from The DHS Program website using the **dhsid** variable. **Dhsid** is a concatenation of the 2-character DHS country code (e.g., **BD** for Bangladesh), the 4-digit year variable (e.g., **2014**), and the 8-digit cluster number (e.g., **00000008**) into a 14-character string (e.g., **BD201400000008**); the content matches the **dhsid** variable in the original DHS files supplying GPS cluster locations.

Case study: using IPUMS DHS contextual variables

As a contribution to the literature and an illustrative example, we consider whether there is an association among exposure to battles, population density, and intimate partner violence with a weapon (IPVW) for ever-partnered women in 13 sub-Saharan African countries.¹⁰ Studies have linked exposure to political conflicts to rates of intrafamilial violence and acceptance of such family violence (Kelly et al. 2018; La Mattina and Shemyakina 2018; Østby 2016). To date, however, such studies have not considered whether relatively low levels of exposure to battles, outside the context of all-out war, have similar influences.

For population density, prior research suggests varying predictions. Greater population density is associated with higher levels of violent crime, including aggravated assault and rape (Bouffard and Muftic 2006). The theoretical explanation is that in low-population-density locations, individuals are more likely to know one another and exercise informal social control to prevent violent crime. In contrast, studies of most forms of IPV show no relationship or a negative relationship with population density. A review of US research discerned no notable differences across urban, suburban, and rural areas in the prevalence of most forms of IPV (Edwards 2015), while Balogum et al. (2012) found lower levels of physical IPV among urban versus rural women in Nigeria. Meanwhile, intimate partner homicide rates—which are likely related to IPVW—are lower in urban areas (Edwards 2015; Gallup-Black 2005). Overall, the IPV literature paints

¹⁰ Burundi 2016, Cameroon 2011, Democratic Republic of Congo 2013, Cote D'Ivoire 2011, Kenya 2014, Malawi 2016, Mali 2012, Mozambique 2011, Nigeria 2013, Rwanda 2014, Zimbabwe 2015, Tanzania 2015, and Zambia 2013. These countries were selected because they had all experienced some low levels of conflict, and the surveys included questions on IPV. Enumeration areas where there is intense conflict, as in parts of the Democratic Republic of Congo, are generally omitted, for interviewer security reasons, from DHS samples.

a different picture than the violent crime literature, suggesting an either a negative or no association between population density and IPVW.

A final consideration is whether the two contextual variables, **popdensity** and **battles**, will interact, with the impact of battles differing depending on the level of population density. Dense urban areas are more likely characterized by high levels of social disorganization (Kunnuji 2016). In such environments, the disruptive impact of battles could be exacerbated, resulting in a positive interaction effect. Contrariwise, because of the typically stronger social ties in rural areas (Bouffard and Muftic 2006), the social disruption of battles may be more intense in low-density areas. If the latter is the case, we will see a negative interaction.

Both IPVW (Fig. 1) and exposure to battles (Fig. 2) are rare occurrences, warranting a pooled sample to conduct the analysis. Using IPUMS DHS, we downloaded a single dataset that included all respondents for each of the 13 samples and the variables relevant to our analysis, harmonized across samples.

DHS surveys for the 13 samples asked women whether their partners had threatened them with a knife, gun, or other type of weapon in the last 12 months (our dependent variable). The key independent variables are whether a woman had been exposed to at least 1 day of battle in the year prior to the survey and population density. We used the **battles** variable, which includes battles within 10 km of each woman's cluster location, to construct exposure to battles in the year preceding the survey. We interpolated values of population density to match survey years using **popdensity_2010** and **popdensity_2015** and logged the resulting values.

We used logistic regression (applying the domestic violence sample weight), including a number of individual- and household-level control variables and

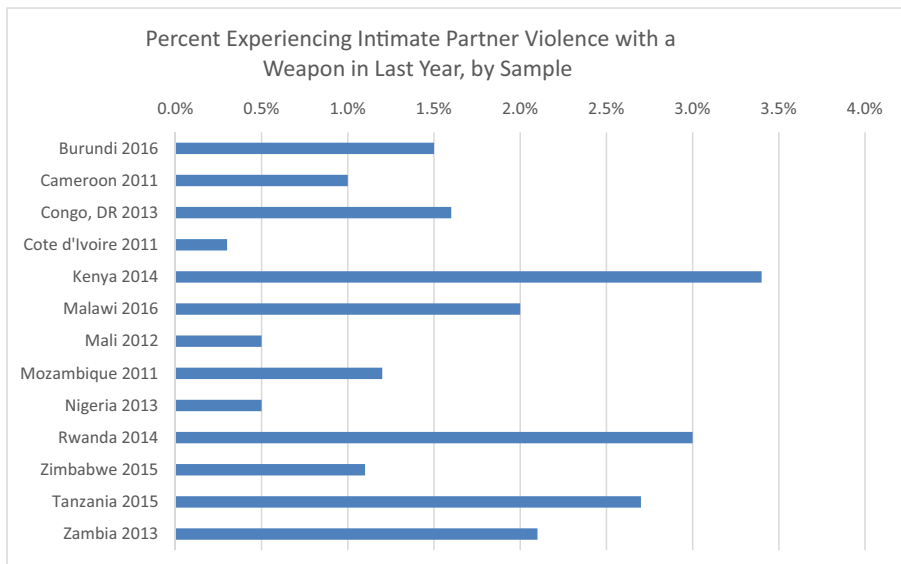


Fig. 1 Percent experiencing intimate partner violence with a weapon in last year, by sample

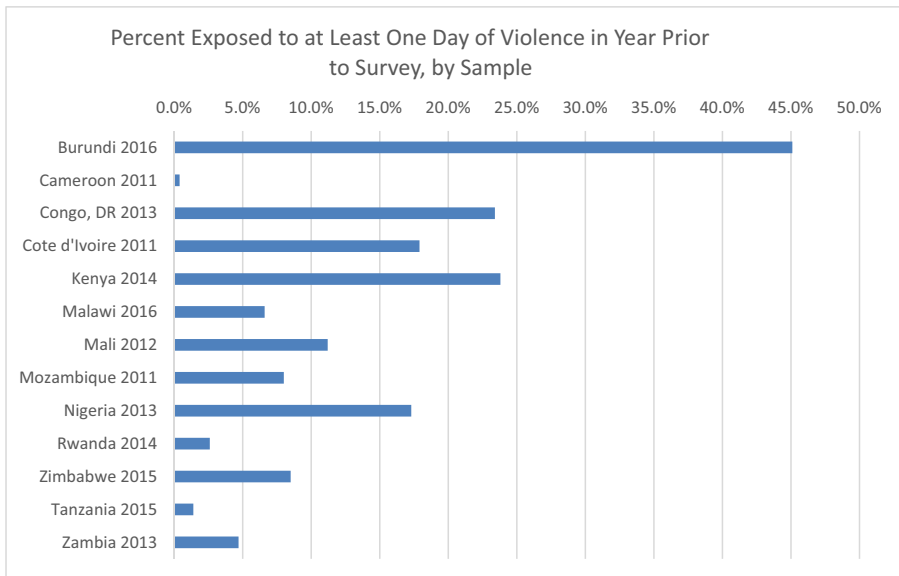


Fig. 2 Percent exposed to at least 1 day of violence in year prior to survey, by sample

country fixed effects. Table 3 reports the results. Model 1 shows the results with the main effects for battles and population density, and Model 2 shows the results with an interaction between battles and population density. Model 1 reveals that battles are associated with an increased probability of IPVW (OR = 1.400; 95% CI = 1.025, 1.905; $p < 0.05$). This is consistent with prior research considering the effect of war on IPV (e.g., Kelly et al. 2018). The model also uncovers a positive association between population density and the risk of IPVW (OR = 1.072; 95% CI = 1.009, 1.140; $p < 0.05$), suggesting that theories of violent crime may apply to IPVW. Model 2 indicates a modest but statistically significant negative interaction between battles and population density (OR = 0.8332; 95% CI = 0.702, 0.989; $p < 0.05$). Among women unexposed to battles, there is little variation in the probability of IPVW across low- and high-population-density areas. Holding their risk of IPVW constant, Fig. 3 shows the marginal probabilities for women exposed to battles, across levels of population density. Among women exposed to battles, IPVW likelihood is significantly greater when population density is lower. This finding suggests that battles are especially disruptive of the dense social ties that often characterize less populated regions.

This preliminary analysis, illuminating previously unstudied relationships between population behavior and the environment, illustrates the ease with which researchers can combine traditional DHS variables with contextual data using IPUMS DHS.

Limitations and future plans

Over time, IPUMS DHS will add more contextual variables. One limitation of the data, however, is that researchers may not find a particular contextual variable of interest

Table 3 Logistic regression predicting experience of intimate partner violence with a weapon in the last year, $N = 210,768$ (weighted)

	Model 1 Log odds (confid. interval)	Model 2 Log odds (confid. interval)
Main effects		
At least 1 day of battles within 10 km of cluster location	0.335* (0.025, 0.645)	1.621** (0.407, 2.833)
Population density, logged	0.070* (0.008, 0.131)	0.095** (0.030, 0.160)
Interaction effect		
At least 1 day of battles within 10 km of cluster location * population density, logged		- 0.182* (- 0.353, - 0.010)
Education		
No education (reference)		
Primary	0.444*** (0.207, 0.680)	0.456*** (0.218, 0.689)
Secondary	0.195 (- 0.113, 0.502)	0.207 (- 0.109, 0.503)
Higher	- 0.474 (- 1.061, 0.113)	- 0.463 (- 1.048, 0.122)
Work		
No paid work (reference)		
Paid in cash	0.375*** (0.187, 0.562)	0.376*** (0.175, 0.553)
Paid in cash and in kind	0.464** (0.200, 0.727)	0.467*** (0.210, 0.737)
Paid in kind only	0.211 (- 0.198, 0.621)	0.190 (- 0.207, 0.623)
Age	- 0.000 (- 0.010, 0.009)	- 0.000 (- 0.009, 0.009)
Wealth		
Poorest	- 0.015 (- 0.253, 0.223)	- 0.004 (- 0.217, 0.263)
Poorer	- 0.063 (- 0.299, 0.174)	- 0.060 (- 0.266, 0.213)
Middle (reference)		
Richer	- 0.163 (- 0.419, 0.094)	- 0.174 (- 0.498, 0.002)
Richest	- 0.494** (- 0.826, - 0.162)	- 0.485*** (- 0.946, - 0.282)
Sample fixed effects (not shown)		
Constant	- 5.062*** (- 5.639, - 4.483)	- 5.300*** (- 5.696, - 4.426)

Asterisks denote statistical significance where * $p < .05$, ** $p < .01$, *** $p < .001$. Confidence intervals are 95%

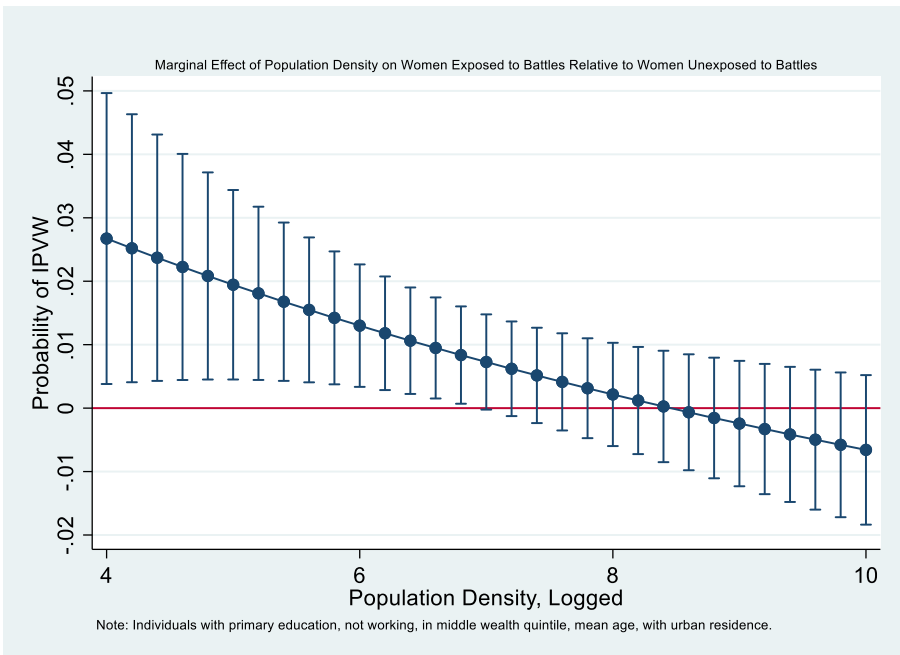


Fig. 3 Marginal effect of population density on women exposed to battles relative to women unexposed to battles

from a preferred source. Likewise, while the IPUMS DHS team has utilized some of the most popular aggregation techniques for spatial and temporal data, there is an infinite range of possibilities. Researchers who prefer other techniques may prefer to link datasets through other means.

A second cautionary note is that researchers should evaluate whether dates when contextual data were collected are reasonable matches for their sample(s) of interest. For example, the non-time-variant agricultural variables (types of crops) derive from data collected in the year 2000. IPUMS DHS leaves it to researchers to investigate the appropriateness of using such data with DHS data collected before 2000 or significantly later. IPUMS DHS gives researchers the flexibility to make these decisions themselves.

Looking ahead, IPUMS DHS has specific plans to increase the suite of contextual variables. Already mentioned is adding summary variables on monthly precipitation and temperature values averaged across decades as baselines for comparison. We will also be adding linking keys that join women through DHS cluster locations to the most temporally proximate international census data from IPUMS-International. This innovation will allow researchers to aggregate census indicators across small-area geography and observe their effects on non-elderly adults' and children's health. Additional plans include incorporating into IPUMS DHS the geospatial covariates, such as night lights, travel times to various locations, and frost occurrences, from the contextual variables calculated by The DHS Program's staff and offered through The DHS Program website (at <http://spatialdata.dhsprogram.com/covariates>). Researchers are encouraged to contact the authors

with additional suggestions of important contextual variables to include in IPUMS DHS.

Conclusion

Social scientists recognize that people's actions and well-being are shaped not only by their individual and familial characteristics but also by community-level characteristics—including transformations such as climate change and geopolitical conflict. While diverse datasets supply georeferenced information on the physical and social environment, locating such information and linking it to other data sources, such as DHS surveys, has been a challenge for geographers and a daunting barrier for researchers from other disciplines. IPUMS DHS makes it easy for researchers to study the effects of the physical environment, population features, and agriculture on health and well-being in LMICS, through a suite of contextual variables. These contextual variables use GPS location data on DHS sample clusters and georeferenced source data from outside the DHS, so researchers can easily study how context shapes health and human behavior. While we have cited research examples, many opportunities remain for understanding health and well-being in LMICS by pairing DHS and contextual data, using IPUMS DHS.

Acknowledgments The authors would like to gratefully acknowledge the support of Sula Sarkar, Greta Gangestad, Zahra Khani, Quinn Heimann, Marya Bakhtiyarova, Dave Van Riper, Kathryn Grace, and Matthew Sobek.

Funding information The authors also appreciate support from the Minnesota Population Center (P2C HD041023), funded through a grant from the Eunice Kennedy Shriver National Institute for Child Health and Human Development (NICHD). IPUMS DHS is supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, grant number R01HD069471, and by USAID and ICF International.

Compliance with ethical standards

This study complies with the current laws of the USA.

References

- Acheson, E. (2015). Predicting disease vector distributions through space and time using environmental and vector control data (doctoral dissertation). Retrieved from Université d'Ottawa/University of Ottawa: https://ruor.uottawa.ca/bitstream/10393/32797/1/Acheson_Emily_2015_thesis.pdf.
- Alemu, Z. A., Ahmed, A. A., Yalaw, A. W., Birhanu, B. S., & Zaitchik, B. F. (2017). Individual and community level factors with a significant role in determining child height-for-age Z score in East Gojjam zone, Amhara regional state, Ethiopia: A multilevel analysis. *Archives of Public Health, 75*(1), 27–40.
- Amoah, B., Giorgi, E., Heyes, D. J., Burren, S., & Diggle, P. J. (2018). Geostatistical modelling of the association between malaria and child growth in Africa. *Int J Health Geogr, 17*(1), 7–19.
- Bakhtiyarova, M., Grace, K., & Nawrotzki, R. J. (2018). Climate, birth weight, and agricultural livelihoods in Kenya and Mali. *Am J Public Health, 108*(S2), S144–S150.
- Balk, D., Storeygard, A., Levy, M., Gaskell, J., Sharma, M., & Flor, R. (2005). Child hunger in the developing world: An analysis of environmental and social correlates. *Food Policy, 30*(5–6), 584–611.
- Balogun, M. O., Owoaje, E. T., & Fawole, O. I. (2012). Intimate partner violence in southwestern Nigeria: Are there rural-urban differences? *Women Health, 52*(7), 627–645.

- Bhatt, S., Weiss, D., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., ... Eckhoff, P. (2015). The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature*, *526*(7572), 207–224.
- Bouffard, L.A. & Muftic, L.R. (2006) The 'rural mystique': Social disorganization and violence beyond urban communities. *Western Criminological Review*, *7*(3).
- Boyle, E.H., King, M.L., & Sobek, M. (2019). IPUMS demographic and health surveys. Version 7 [dataset]. *Minnesota Population Center and ICF International*. <https://doi.org/10.18128/D080.V7>
- Brown, M. E., Grace, K., Shively, G., Johnson, K. B., & Carroll, M. (2014). Using satellite remote sensing and household survey data to assess human health and nutrition response to environmental change. *Popul Environ*, *36*(1), 48–72.
- Burgert, C., Colston, J., Roy, T., & Zachary, B. (2013) Georeferenced displacement procedures and georeferenced data release policy for the Demographic and Health Surveys. DHS spatial analysis reports 7. ICF.
- Butz, W. P., & Torrey, B. B. (2006). Some frontiers in social science. *Science*, *312*(5782), 1898–1900.
- Carranza, E. (2014). Soil endowments, female labor force participation, and the demographic deficit of women in India. *Am Econ J Appl Econ*, *6*(4), 197–225.
- Center for International Earth Science Information Network - CIESIN - Columbia University. (2016). Documentation for the gridded population of the world, version 4 (GPWv4). Retrieved from Palisades NY: <http://beta.sedac.ciesin.columbia.edu/data/collection/gpw-v4/documentation>.
- Corsi, D. J., Neuman, M., Finlay, J. E., & Subramanian, S. (2012). Demographic and Health Surveys: A profile. *Int J Epidemiol*, *41*(6), 1602–1613.
- Cuadros, D. F., Branscum, A. J., & Crowley, P. H. (2011a). HIV–malaria co-infection: Effects of malaria on the prevalence of HIV in East sub-Saharan Africa. *Int J Epidemiol*, *40*(4), 931–939.
- Cuadros, D. F., Branscum, A. J., & García-Ramos, G. (2011b). No evidence of association between HIV-1 and malaria in populations with low HIV-1 prevalence. *PLoS One*, *6*(8), e23458.
- Cunningham, D. E., Gleditsch, K. S., & Salehyan, I. (2009). It takes two: A dyadic analysis of civil war duration and outcome. *J Confl Resolut*, *53*, 570–597.
- Darrouzet-Nardi, A. F. (2016). Nonviolent civil insecurity is negatively associated with subsequent height-for-age in children aged < 5 y born between 1998 and 2014 in rural areas of Africa, *2. Am J Clin Nutr*, *105*(2), 485–493.
- Davenport, F., Grace, K., Funk, C., & Shukla, S. (2017). Child health outcomes in sub-Saharan Africa: A comparison of changes in climate and socio-economic factors. *Glob Environ Chang*, *46*, 72–87.
- Delavande, A., & Menezes Cordeiro, R. (2012). *Violent conflicts and risky sexual behavior in Uganda (MICROCON Research Working Paper 60)*. Brighton: MICROCON.
- Didan, K. (2015). MOD13Q1 MODIS/Terra vegetation indices 16-Day L3 global 250m SIN grid V006. Retrieved from <https://lpdaac.usgs.gov/node/844>.
- Edwards, K. M. (2015). Intimate partner violence and the rural–urban–suburban divide: Myth or reality? A critical review of the literature. *Trauma, Violence, & Abuse*, *16*(3), 359–373.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... Hoell, A. (2015). The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Scientific Data*, *2*, 150066.
- Gallup-Black, A. (2005). Twenty years of rural and urban trends in family and intimate partner homicide: Does place matter? *Homicide Stud*, *9*(2), 149–173.
- Grace, K., & Nagle, N. N. (2015). Using high-resolution remotely sensed data to examine the relationship between agriculture and fertility in Mali. *Prof Geogr*, *67*(4), 641–654.
- Grace, K., Davenport, F., Funk, C., & Lerner, A. M. (2012). Child malnutrition and climate in sub-Saharan Africa: An analysis of recent trends in Kenya. *Appl Geogr*, *35*(1–2), 405–413.
- Grace, K., Brown, M., & McNally, A. (2014). Examining the link between food prices and food insecurity: A multi-level analysis of maize price and birthweight in Kenya. *Food Policy*, *46*, 56–65.
- Grace, K., Davenport, F., Hanson, H., Funk, C., & Shukla, S. (2015). Linking climate change and health outcomes: Examining the relationship between temperature, precipitation and birth weight in Africa. *Glob Environ Chang*, *35*, 125–137.
- Grace, K., Frederick, L., Brown, M. E., Boukerrou, L., & Lloyd, B. (2017). Investigating important interactions between water and food security for child health in Burkina Faso. *Popul Environ*, *39*(1), 26–46.
- Grace, K., Nagle, N. N., Burgert-Brucker, C. R., Rutzick, S., Van Riper, D. C., Dontamsetti, T., & Croft, T. (2019). Integrating environmental context into DHS analysis while protecting participant confidentiality: A new remote sensing method. *Popul Dev Rev*, *45*(1), 197–218.

- Headey, D. D., & Hoddinott, J. (2016). Agriculture, nutrition and the green revolution in Bangladesh. *Agric Syst*, 149, 122–131.
- Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, 12(2), e0169748.
- Hirvonen, K., & Hoddinott, J. (2017). Agricultural production and children's diets: Evidence from rural Ethiopia. *Agric Econ*, 48(4), 469–480.
- Hyland, M., & Russ, J. (2019). Water as destiny—the long-term impacts of drought in sub-Saharan Africa. *World Dev*, 115, 30–45.
- Ickowitz, A., Powell, B., Salim, M. A., & Sunderland, T. C. (2014). Dietary quality and tree cover in Africa. *Glob Environ Chang*, 24, 287–294.
- Jankowska, M. M., Lopez-Carr, D., Funk, C., Husak, G. J., & Chafe, Z. A. (2012). Climate change and human health: Spatial modeling of water availability, malnutrition, and livelihoods in Mali, Africa. *Appl Geogr*, 33, 4–15.
- Johnson, K., & Brown, M. E. (2014). Environmental risk factors and child nutritional status and survival in a context of climate variability and change. *Appl Geogr*, 54, 209–221.
- Jones, A. D., Shrinivas, A., & Bezner-Kerr, R. (2014). Farm production diversity is associated with greater household dietary diversity in Malawi: Findings from nationally representative data. *Food Policy*, 46, 1–12.
- Joy, E. J., Bradley, M. R., Young, S. D., Black, C. R., Chilimba, A. D., Ander, E. L., ... Watts, M. J. (2015). Soil type influences crop mineral composition in Malawi. *Science of the Total Environment*, 505, 587–595.
- Kazembe, L. N., Appleton, C. C., & Kleinschmidt, I. (2007). Spatial analysis of the relationship between early childhood mortality and malaria endemicity in Malawi. *Geospat Health*, 2(1), 41–50.
- Kelly, J. T., Colantuoni, E., Robinson, C., & Decker, M. R. (2018). From the battlefield to the bedroom: A multilevel analysis of the links between political conflict and intimate partner violence in Liberia. *BMJ Glob Health*, 3(2), e000668.
- Kunnuji, M. O. (2016). Population density and armed robbery in Nigeria: An analysis of variation across states. *African Journal of Criminology and Justice Studies*, 9(1), 62–73.
- La Mattina, G., & Shemyakina, O. (2018). Domestic violence and childhood exposure to armed conflict: Attitudes and experiences (working paper no. 255). Retrieved from households in conflict network website: <http://www.hicn.org/wordpress/wp-content/uploads/2012/06/HiCN-WP255.pdf>.
- León, F. R. (2012). The latitudinal tilts of wealth and education in Peru: Testing them, explaining them, and reflecting on them. *Economía*, 35(70), 60.
- Lillepold, K., Aimone, A., Keino, S., & Braitstein, P. (2018). Assessing the effect of drought severity on height-for-age z-score in Kenyan children: A secondary analysis. *The Lancet Planetary Health*, 2, S13.
- Liu, J., Fritz, S., Van Wesenbeeck, C., Fuchs, M., You, L., Obersteiner, M., & Yang, H. (2008). A spatially explicit assessment of current and future hotspots of hunger in sub-Saharan Africa in the context of global change. *Glob Planet Chang*, 64(3–4), 222–235.
- Livelihoods (2020, March 26). Retrieved from <https://fews.net/livelihoods>.
- Mabaso, M. (1998). Environmental factors influencing the distribution hookworm infection in KwaZulu-Natal, South Africa (Doctoral Dissertation). Retrieved from https://www.researchgate.net/profile/Musawenkosi_Mabaso/publication/266294343_ENVIRONMENTAL_FACTORS_INFLUENCING_THE_DISTRIBUTION_HOOKWORM_INFECTION_IN_KWAZULU_NATAL_SOUTH_AFRICA/links/54bca46c0cf29e0b04c169b.pdf.
- Mabaso, M., Appleton, C., Hughes, J., & Gouws, E. (2003). The effect of soil type and climate on hookworm (*Necator americanus*) distribution in KwaZulu Natal, South Africa. *Tropical Med Int Health*, 8(8), 722–727.
- MacVicar, S. (2016). The impacts of meteorological exposures on perinatal health: A mixed methods study of Kanungu District, Uganda (Doctoral Dissertation, McGill University Libraries).
- Malcomb, D. W., Weaver, E. A., & Krakowka, A. R. (2014). Vulnerability modeling for sub-Saharan Africa: An operationalized approach in Malawi. *Appl Geogr*, 48, 17–30.
- Matthews, S. A. (2008). Advanced spatial analysis training program for population scientists—Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) R25 proposal. Population Research Institute, Pennsylvania State University.
- Mohammed, M. A. (2016). Patterns of children's work during schooling and its association with school attendance in coffee and non-coffee picking seasons in intense coffee growing regions in Sidama Zone, Southern Ethiopia: A mixed-methods study (Master's Thesis). Retrieved from <http://hdl.handle.net/10642/3298>.

- Monfreda, C., Ramankutty, N., & Foley, J. A. (2008). Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochemical Cycles*, 22(1).
- Mulmi, P., Block, S. A., Shively, G. E., & Masters, W. A. (2016). Climatic conditions and child height: Sex-specific vulnerability and the protective effects of sanitation and food markets in Nepal. *Economics & Human Biology*, 23, 63–75.
- National Geospatial-Intelligence Agency. (2012). World geodetic system 1984 (WGS84). Retrieved from <https://web.archive.org/web/20120402143802/https://www.nga.mil/ProductsServices/GeodesyandGeophysics/WorldGeodeticSystem/Pages/default.aspx>.
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., et al. (2001). Terrestrial ecoregions of the world: A new map of life on earth. *BioScience*, 51(11), 933–938.
- Østby, G. (2016). Violence begets violence: Armed conflict and domestic sexual violence in sub-Saharan Africa. Paper presented at the workshop on sexual violence and armed conflict: New Research Frontiers held at the Harvard Kennedy School, Harvard University.
- Perez-Heydrich, C., Warren, J. L., Burgert, C. R., & Emich, M. E. (2016). Influence of demographic and health survey point displacements on raster-based analyses. *Spatial Demography*, 4(2), 135–153.
- Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing ACLED: An armed conflict location and event dataset: Special data feature. *J Peace Res*, 47(5), 651–660.
- Ramankutty, N., Evan, A. T., Monfreda, C., & Foley, J. A. (2008). Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles*, 22(1).
- Rohde, J. T. (2014). The Relationship between access to healthcare and civil conflict (Undergraduate Honors Thesis). Retrieved from <https://scholarworks.wm.edu/cgi/viewcontent.cgi?article=1099&context=honorsthesis>.
- Rosenberg, A. M., Maluccio, J. A., Harris, J., Mwanamwenge, M., Nguyen, P. H., Tembo, G., & Rawat, R. (2018). Nutrition-sensitive agricultural interventions, agricultural diversity, food access and child dietary diversity: Evidence from rural Zambia. *Food Policy*, 80, 10–23.
- Sheffield, J., Goteti, G., & Wood, E. F. (2006). Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling. *J Climate*, 19(13), 3088–3111 <http://hydrology.princeton.edu/home.php> <http://hydrology.princeton.edu/home.php>.
- Slavchevska, V. (2015). Agricultural production and the nutritional status of family members in Tanzania. *J Dev Stud*, 51(8), 1016–1033.
- Vepa, S. S., Viswanathan, B., Parasar, R., & Bhavani, R. (2016). Child underweight, land productivity and public services: A district-level analysis for India (working paper no. 6). Retrieved from <http://www.lansasouthasia.org/content/child-underweight-land-productivity-and-public-services-district-level-analysis-india>.
- Wayoro, D. (2018). Impact of armed conflicts on child welfare in Côte d’Ivoire. Paper presented at the meeting of the American Economic Association, Philadelphia, PA.
- Weir, J. & Herring, D. (2013) Measuring vegetation (NDVI and EVI). NASA earth observatory library. Retrieved from <http://earthobservatory.nasa.gov/Features/MeasuringVegetation/>.
- Woodruff, B. A., Wirth, J. P., Bailes, A., Matji, J., Timmer, A., & Rohner, F. (2017). Determinants of stunting reduction in Ethiopia 2000–2011. *Maternal & Child Nutrition*, 13(2), e12307.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.