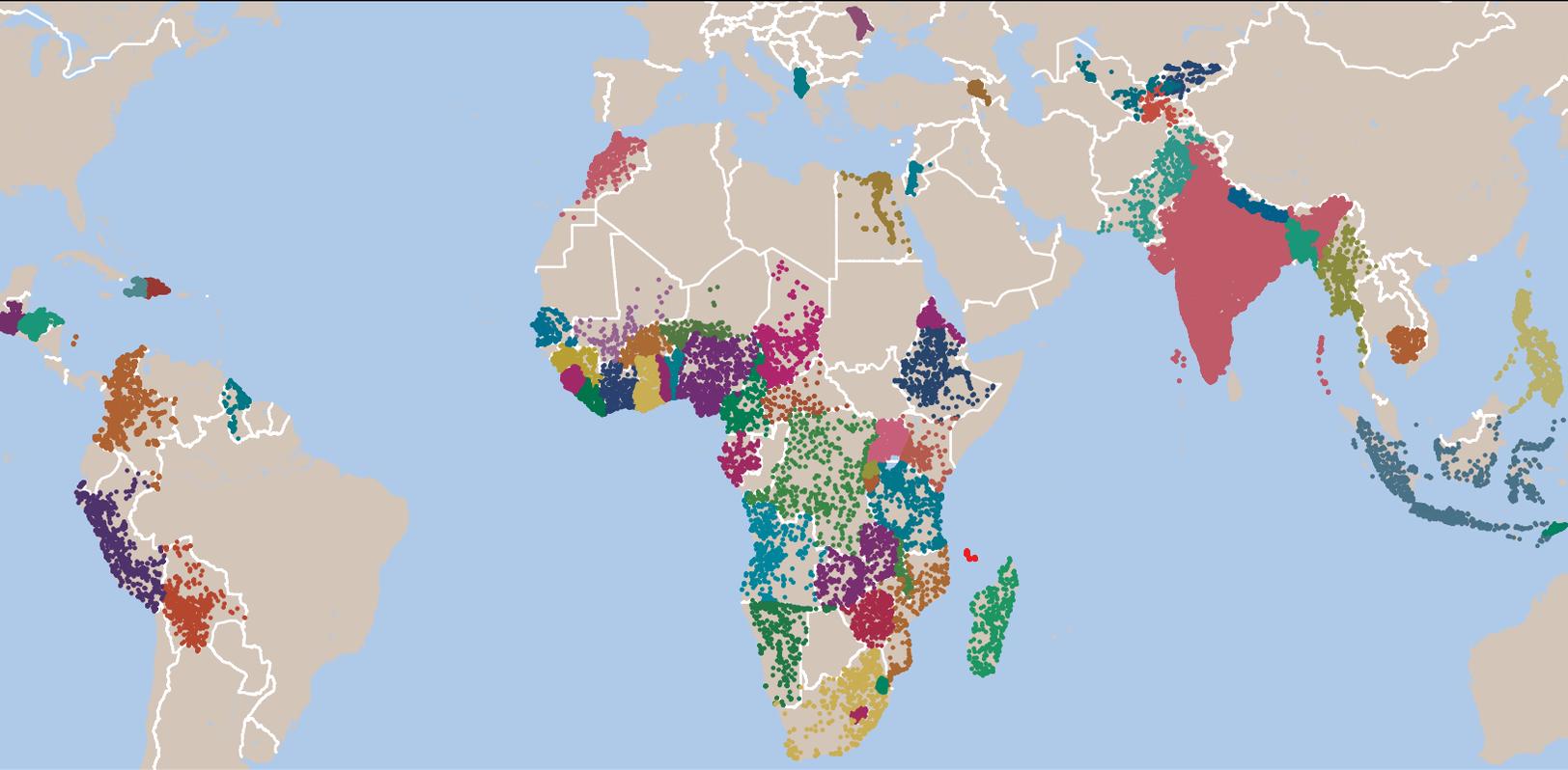




USAID
FROM THE AMERICAN PEOPLE

PREDICTING HIV/AIDS AT SUBNATIONAL LEVELS USING DHS COVARIATES RELATED TO HIV

DHS SPATIAL ANALYSIS REPORTS 18



March 2020

This publication was produced for review by the United States Agency for International Development (USAID). The report was prepared by Benjamin K. Mayala, Samir Bhatt, and Peter Gething.

DHS Spatial Analysis Reports No. 18

**Predicting HIV/AIDS at Subnational Levels
using DHS Covariates related to HIV**

Benjamin K. Mayala

Samir Bhatt

Peter Gething

The DHS Program

ICF

Rockville, Maryland, USA

March 2020

Corresponding author: Benjamin K. Mayala, The DHS Program, 530 Gaither Road, Suite 500, Rockville, MD 20850, USA; phone: 301-572-0507; fax: 301-407-6501; email: Ben.Mayala@icf.com

Acknowledgments: The authors would like to acknowledge Trinadh Dontamsetti, Bradley Janocha, Joy Fishel, and Trevor Croft for their assistance and guidance in developing this report.

Editor: Diane Stoy

Document Production: Joan Wardell

This study was implemented 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 DHS 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:

Mayala, Benjamin. K., Samir Bhatt, and Peter Gething. 2020. *Predicting HIV/AIDS at Subnational Levels using DHS Covariates related to HIV*. DHS Spatial Analysis Reports No. 18. Rockville, Maryland, USA: ICF.

CONTENTS

TABLES	v
FIGURES	vii
PREFACE	xi
ABSTRACT	xiii
ACRONYMS AND ABBREVIATIONS	xv
1 BACKGROUND AND OBJECTIVES	1
1.1 Background	1
1.2 Objective	2
2 DATA PREPARATION	5
2.1 Indicators	5
2.2 Data Plots	6
3 METHODS	15
3.1 Modelling Rationale	15
3.2 Modelling Limitations	16
3.3 Modelling Framework	16
4 RESULTS	19
4.1 Correlation Plots	19
4.2 HIV Prevalence Maps and Confidence Interval Plots	24
5 CONCLUSION	29
REFERENCES	31

TABLES

Table 1	Description of DHS indicators used in the study	5
---------	---	---

FIGURES

Figure 1	Burundi 2016: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.	6
Figure 2	DRC 2013: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.	7
Figure 3	Ethiopia 2011: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.	8
Figure 4	Ethiopia 2016: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.	9
Figure 5	Ghana 2014: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.	10
Figure 6	Mali 2014: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.	11
Figure 7	Rwanda 2015: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.	12

Figure 8	South Africa 2016: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.	13
Figure 9	Correlation plot for Burundi 2016. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.	19
Figure 10	Correlation plot for DRC 2013. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.	20
Figure 11	Correlation plot for Ethiopia 2011. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.	20
Figure 12	Correlation plot for Ethiopia 2016. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.	21
Figure 13	Correlation plot for Ghana 2014. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.	21
Figure 14	Correlation plot for Mali 2012. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.	22
Figure 15	Correlation plot for Rwanda 2015. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.	22
Figure 16	Correlation plot for South Africa 2016. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.	23

Figure 17	Burundi 2016. Prevalence map (left) and uncertainty (right).....	24
Figure 18	DRC 2016. Prevalence map (left) and uncertainty (right).....	25
Figure 19	Ethiopia 2011. Prevalence map (left) and uncertainty (right).	25
Figure 20	Ethiopia 2016. Prevalence map (left) and uncertainty (right).	25
Figure 21	Ghana 2014. Prevalence map (left) and uncertainty (right).....	26
Figure 22	Mali 2012. Prevalence map (left) and uncertainty (right).....	26
Figure 23	Rwanda 2015. Prevalence map (left) and uncertainty (right).	27
Figure 24	South Africa 2016. Prevalence map (left) and uncertainty (right).....	27

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 other DHS reports and respond to the increasing interest in a spatial perspective on demographic and health data. The principal objectives of all the 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 DHS 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

Accelerating the scale-up of HIV prevention and treatment approaches will limit the epidemic to more manageable levels and enable countries to move toward the elimination phase. In this context, the World Health Organization produced new treatment guidelines and UNAIDS set the 90-90-90 targets in which by 2020, 90% of all people living with HIV will know their HIV status, 90% of all people with diagnosed HIV infection will receive sustained antiretroviral therapy, and 90% of all people receiving antiretroviral therapy will be virally suppressed. Most household surveys such as the Demographic Health Surveys and the AIDS Indicator Surveys are designed to provide reliable estimates of survey indicators primarily at the national level, as well as the first subnational administrative level. To better address the need for fine spatial and lower level (district) estimates, geospatial modelling techniques that can leverage existing survey data, spatial relationships between survey clusters, and relationships with geospatial covariates have become increasingly popular in mapping key development indicators at high spatial resolutions.

In this report, we use a multitask Gaussian process (GP), in which we use aggregated spatial coordinates and fit multiple Gaussian processes to HIV prevalence and other indicators such as women who reported condom use at last high-risk sex with a non-cohabiting, non-marital partner or the number of partners in lifetime for women. However, instead of using these Gaussian processes independently, we also model the cross correlations between all indicators. This methodology allows for all indicators to inform each other and to increase the predictive power of the model. In addition to prediction, the joint multitask framework also allows for a unified treatment of uncertainty that provides robust uncertainty intervals.

For the first time, the proposed approach uses indicators that are actually relevant to predicting HIV prevalence. We also utilize a Bayesian framework and therefore have a robust treatment of uncertainty. We foresee many applications of this approach to other diseases in the future.

ACRONYMS AND ABBREVIATIONS

ADMIN 1	first subnational administrative level
ADMIN 2	second subnational administrative level
AIS	AIDS Indicator Survey
ART	antiretroviral therapy
CAR	condition autoregressive models
DHS	Demographic and Health Survey
GMRF	Gaussian Markov random field
GP	Gaussian process
HIV	human immunodeficiency virus
MBG	model-based geostatistics
PLHIV	people living with HIV
SAE	small area estimation
SDG	Sustainable Development Goal
STI	sexual transmitted infection
WHO	World Health Organization

1 BACKGROUND AND OBJECTIVES

1.1 Background

Human immunodeficiency virus (HIV) is one of the most devastating diseases of the modern age and has resulted in the death of over 35 million individuals since the epidemic began (WHO 2019). An estimated 36.9 million people were living with HIV in 2017, with the vast majority of these individuals concentrated in sub-Saharan Africa (WHO 2019). Provision of antiretroviral therapy (ART) has enabled people living with HIV (PLHIV) to attain life expectancies similar to those of HIV-negative persons (Nakagawa, May, and Phillips 2013; Bor et al. 2013).

Accelerating the scale-up of HIV prevention and treatment approaches will limit the epidemic to more manageable levels and enable countries to move towards the elimination phase. In this context, the World Health Organization (WHO) produced new treatment guidelines and UNAIDS set the 90-90-90 targets in which by 2020, 90% of all people living with HIV will know their HIV status, 90% of all people with diagnosed HIV infection will receive sustained ART, and 90% of all people receiving ART will be virally suppressed. Many challenges in achieving the 90-90-90 targets center on facilitating data collection, assessing knowledge gaps, and planning treatment strategies, which are all fundamentally underpinned by the ability to determine fine-scale subnational patterns of HIV burden and treatment.

Numerous studies have shown the potential benefits of adjusting HIV programs to focus on the populations and locations with the highest incidence and treatment needs (Hallett et al. 2016; Meyer-Rath et al. 2018). The wide variation in the geographic spread of HIV makes it inefficient to implement the same set of programs uniformly throughout a country (Cuadros and Abu-Raddad 2014; Tiwari et al. 2006; Zulu, Kalipeni, and Johannes 2014). The limited availability of reliable HIV data at subnational levels also makes it difficult for programs to target resources to the locations most in need.

Most household surveys such as the Demographic Health Surveys (DHS) and the AIDS Indicator Surveys (AIS) are designed to provide reliable estimates of survey indicators primarily at the national level as well as the first subnational administrative level (ADMIN 1). Since national level estimates are more useful for comparing nations and aggregating data across large world regions, their natural audience includes international policy makers and donors (Li et al. 2019). Further, the ADMIN 1 analysis does not provide comprehensive estimates at lower levels such as the second subnational administrative level (ADMIN 2), where health programs are designed and implemented (Mayala et al. 2019). The Global Fund to Fight AIDS, Tuberculosis, and Malaria (Global Fund) recognizes the need for data that illustrate the impact of health programs. The Global Fund's 2017-2022 Strategy, "Investing to End Epidemics,"¹ includes an operational objective to "Strengthen data systems for health and countries' capacities for analysis and use." The Strategy notes that fine, spatial, and lower level estimates are essential for good decision-making and are a prerequisite for the success and long-term impact of HIV and other health programs.

To better address the need for fine spatial and lower level estimates, there are two existing options: either scaling up the nationally representative survey data collection process by increasing the sample size, or by

¹ https://www.theglobalfund.org/media/1176/bm35_02-theglobalfundstrategy2017-2022investingtoendepidemics_report_en.pdf

using the geospatial modelling (Bayesian model-based geostatistics - MBG) approach. The approach with increasing sample size is costly and time-consuming. Instead, geospatial modelling techniques that can leverage existing survey data, spatial relationships between survey clusters, and relationships with geospatial covariates have become increasingly popular in mapping key development indicators at high spatial resolutions (Mayala et al. 2019; Utazi et al. 2018).

The MBG approach is increasingly recognized as an excellent geostatistical analysis method for addressing uncertainty in the model estimates (predictions) and for being flexible and capable of handling missing data (Cressie and Wikle 2011). This approach has been widely used to predict and map various indicators such as those described in Spatial Analysis Reports 11 and 17 (Gething et al. 2015; Mayala et al. 2019), poverty (Steele et al. 2017), and malaria (Gething, et al. 2011; Gosoni, Vountsou, Sogoba, 2006; Gosoni, Veta, and Vounatsou 2011; Gosoni et al. 2012; Hay et al. 2009; Kazembe et al. 2006; Raso et al. 2012; Riedel et al. 2011). In these studies, climatic/environmental and socioeconomic data layers that are thought to influence the indicators are used to explain some variations in prevalence across different areas. This information can aid our understanding of the relationships between the indicator and the influence of climatic/environmental and socioeconomic factors (Noor et al. 2009).

Previous studies of mapping subnational estimation of HIV prevalence have focused on several approaches. Ordinary (simple) kriging has been used to model HIV prevalence and identify the vulnerable population at high risk of infection in Tanzania and South Africa (Cuadros et al. 2018; Kleinschmidt et al. 2007). PreVR package (Larmarange et al. 2011)—a kernel density estimation—has been used to estimate subnational level HIV prevalence from DHS data surveys in selected countries in sub-Saharan Africa (Coburn, Okano, and Blower 2017; Larmarange and Bendaud 2014). PreVR does not incorporate the uncertainty of estimation from the DHS and cannot produce quantitative estimates of uncertainty (Gutierrez et al. 2019). Fully Bayesian geostatistical models have been used to estimate subnational HIV prevalence from individual-level health survey data and a range of geospatial covariates, including those not related to HIV/AIDS (Dwyer-Lindgren et al. 2019; Kandala et al. 2012; Ngesa, Mwambi, and Achia 2014; Niragire et al. 2015; Okango, Mwambi, and Ngesa 2016). Unlike the ordinary kriging and kernel density (in PreVR), those Bayesian models accommodate the uncertainty in the data and can produce probability intervals on the estimates.

1.2 Objective

In this report, we introduce a novel framework for the subnational estimation of HIV prevalence using DHS data. For the first time, we utilize a multi-task Gaussian process (GP) to leverage data collected in the survey along with HIV-related indicators such as sexually transmitted infection (STI), condom use, and sex partners. Our model produces estimates of HIV prevalence at ADMIN 2 levels, along with uncertainty. We present the rationale for using this approach, discuss why it is superior to previously used approaches, and describe the methods in detail.

This approach aligns well with the Global Fund’s “Framework for Data Use for Action and Improvement at Country Level.”² The data generated from this modelling would enable in-country partners at the national, subnational and community levels to make better decisions, drive program performance and

² https://www.theglobalfund.org/media/8362/me_datauseforactionandimprovement_framework_en.pdf?u=637001819710000000

outcomes, and achieve the intended impact. This approach is unique because it will: (1) restrict model data input to indicators that are relevant to HIV only, which would avoid over-prediction (other studies have used non-HIV related indicators); (2) restrict the model framework to individual, high-impact Global Fund-supported countries to produce robust models, which are relevant to a country's program planning processes, including the upcoming 2020 grant proposal process.

2 DATA PREPARATION

2.1 Indicators

Table 1 describes the indicators analyzed in this study. These were extracted from nine national DHS surveys, which include Burundi 2017, Democratic Republic of the Congo (DRC) 2014, Ethiopia 2011, Ethiopia 2016, Mali 2014, Mozambique 2015, Rwanda 2015 and South Africa 2016. We considered data only from those surveys that had described the total number of individuals examined, the proportion of positive cases, and the coordinates of their geographical locations.

Table 1 Description of DHS indicators used in the study

Indicator	Definition
Current marital status	Percentage currently married or living in union
Condom use	Percentage of condom use at last high-risk sex with a non-cohabiting, non-marital partner
Sexual intercourse	Percentage of those who ever had sexual intercourse between age 15-24
Sex partners	Number of sex partners in lifetime, calculated by grouping number of sexual partners in lifetime into (<5, >5).
Partners live elsewhere	Percentage of women (men) whose partners live elsewhere
STI	Percentage reporting an STI
STI symptoms	Percentage reporting an STI and symptoms
HIV test	Percentage of who tested HIV positive

2.2 Data Plots

We plot the raw data for all indicators, for all surveys. The axis was not standardized to ensure that data heterogeneity was visible.

Figure 1 Burundi 2016: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.

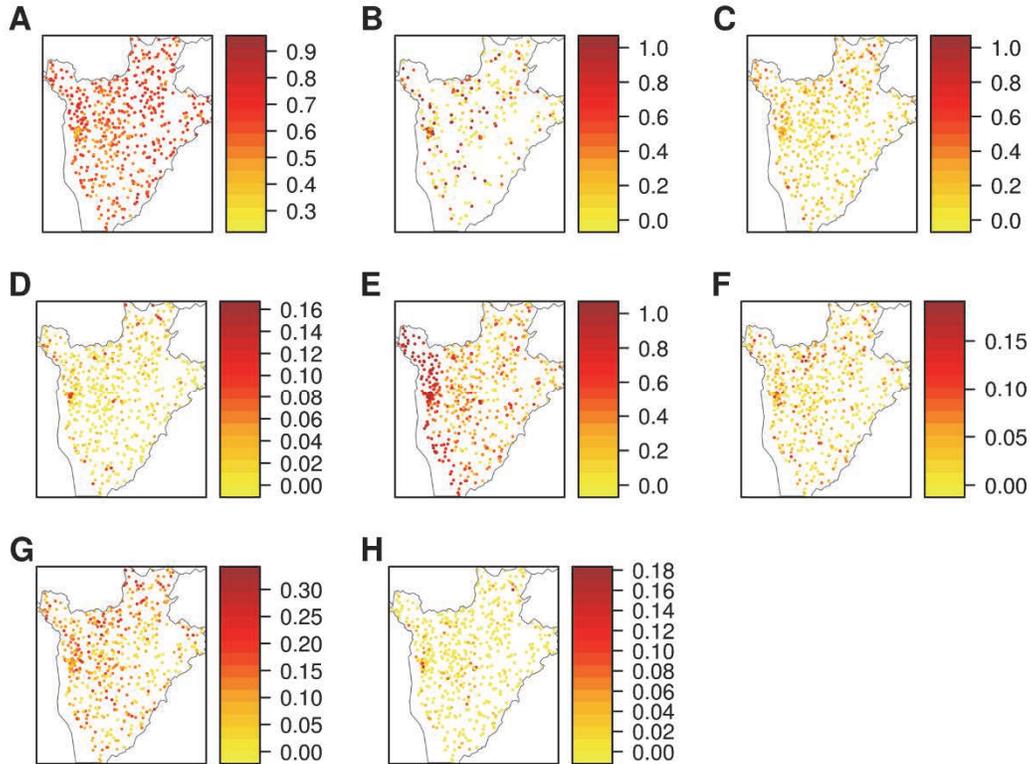


Figure 2 DRC 2013: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.

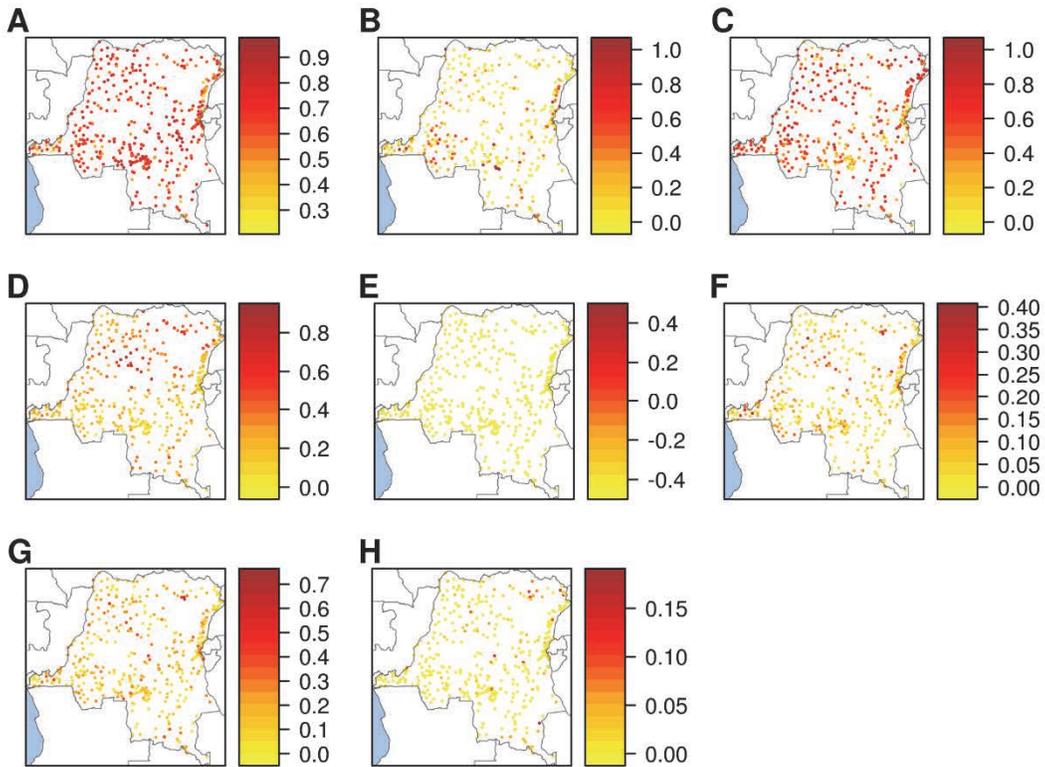


Figure 3 Ethiopia 2011: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.

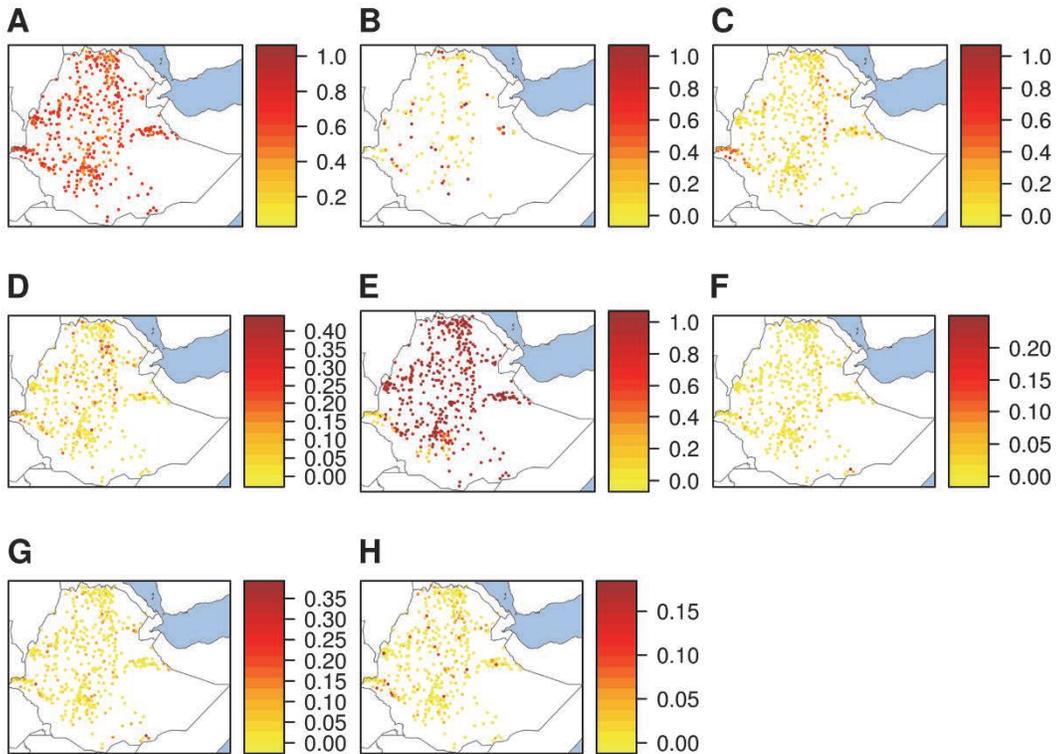


Figure 4 Ethiopia 2016: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.

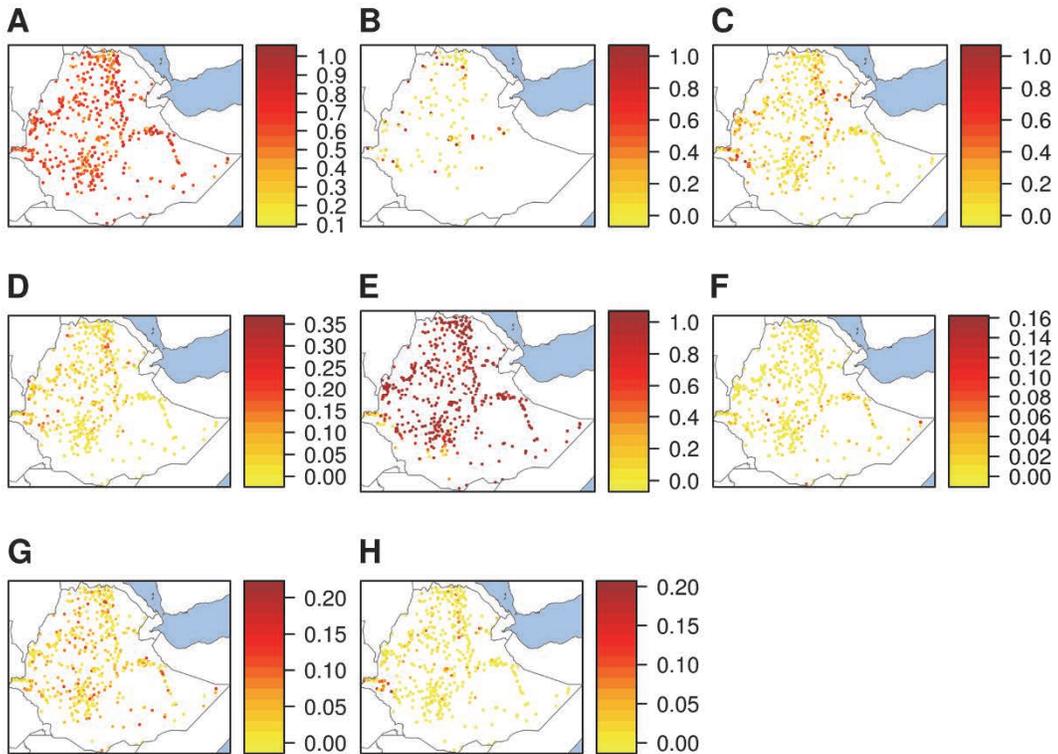


Figure 5 Ghana 2014: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.

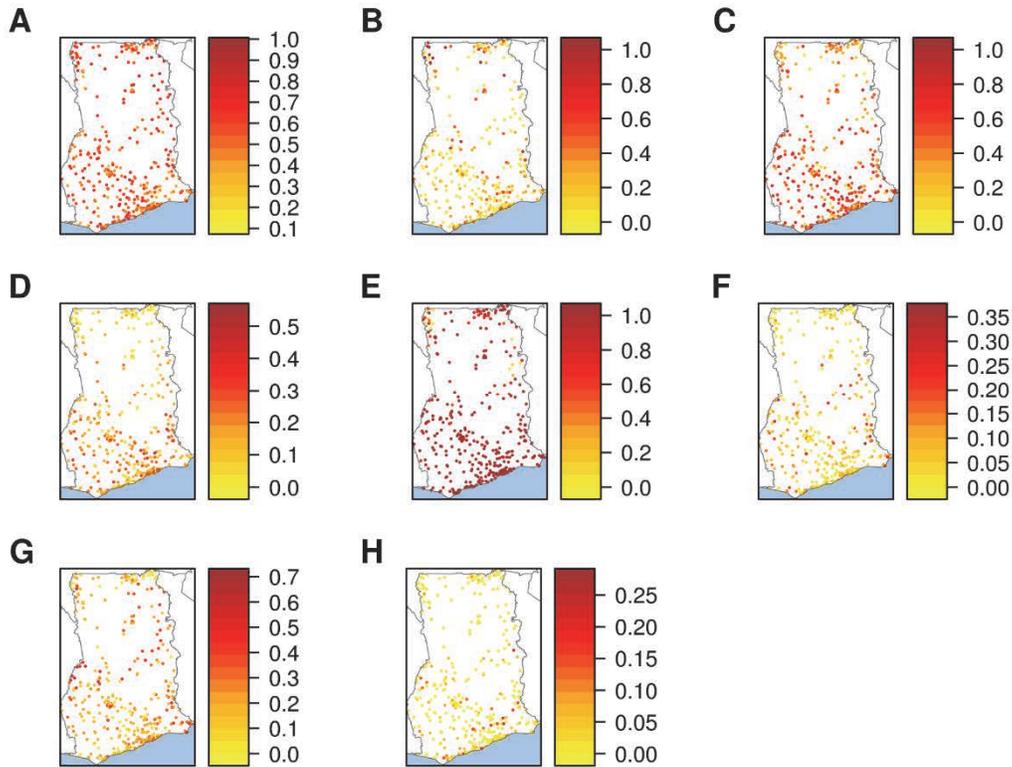


Figure 6 Mali 2014: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.

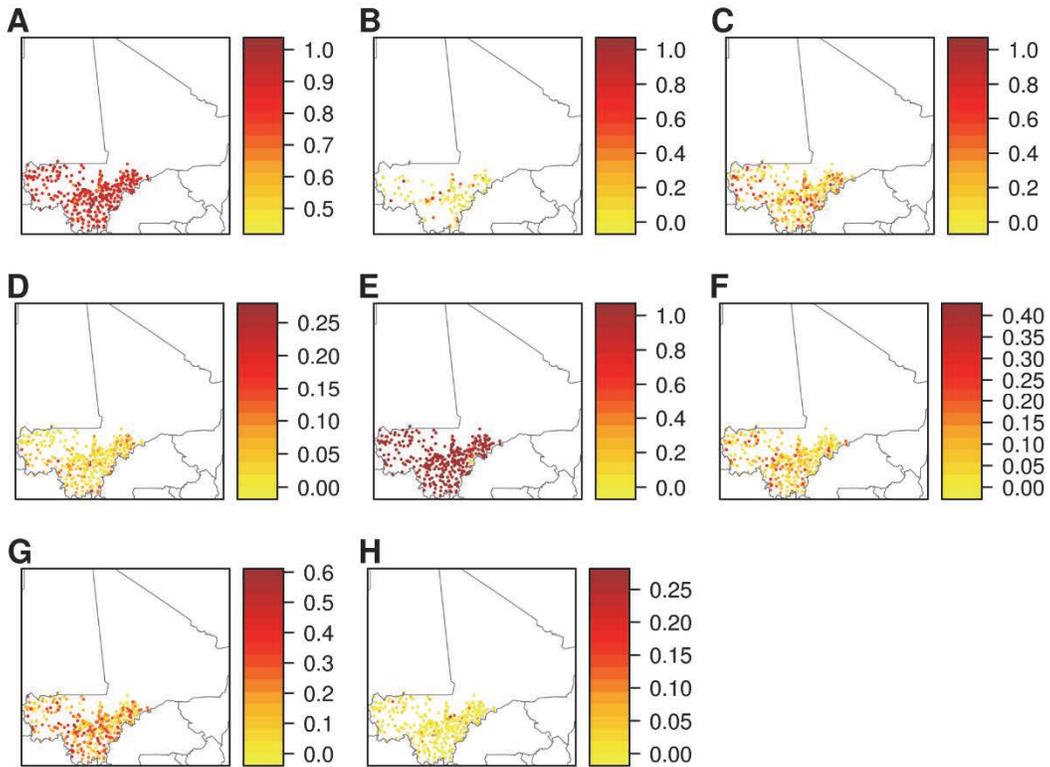


Figure 7 Rwanda 2015: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.

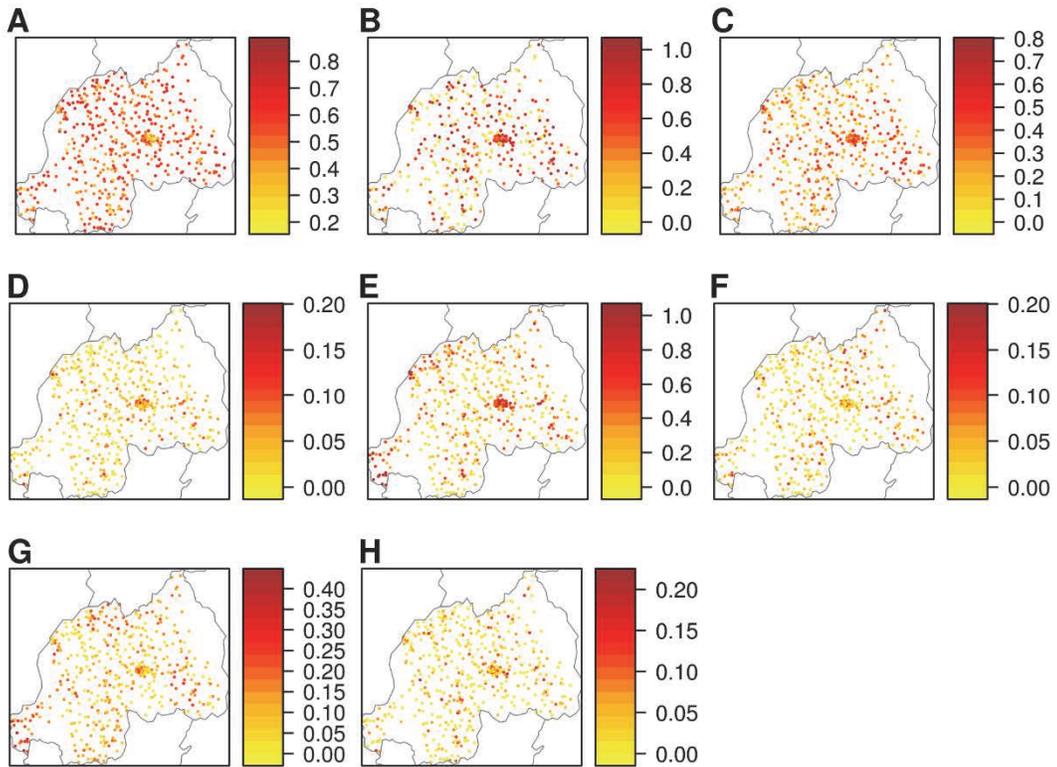
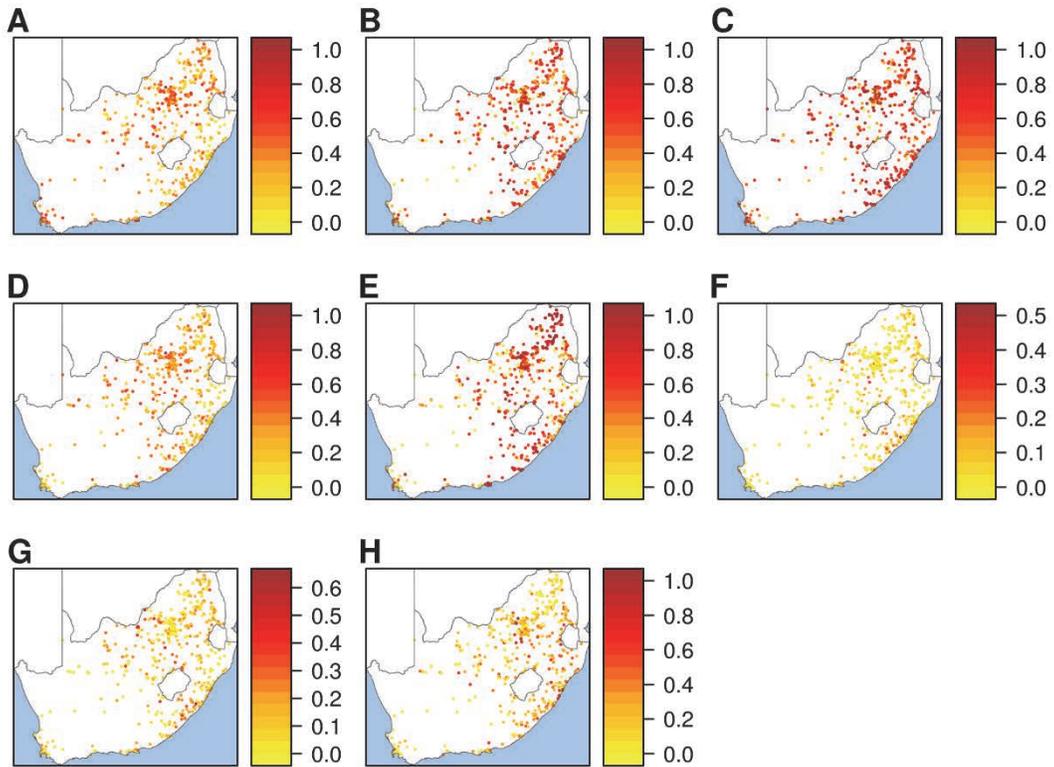


Figure 8 South Africa 2016: Plot of raw data prevalence ratios for (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity.



3 METHODS

3.1 Modelling Rationale

Geostatistical modelling has been used extensively to model survey data (Dwyer-Lindgren et al. 2019), and the central methodology is Gaussian process (GP) regression. In a generic GP regression model, the observed variation in HIV prevalence can broadly be explained by one of three components:

1. Sampling and random error, which can often be large given the small sample sizes at individual clusters, is represented with a standard sampling model (the variance of a Gaussian likelihood).
2. Some non-sampling variation can often be explained using fixed effects or covariates—whereby a multivariate regression relationship is defined on the HIV prevalence via explanatory covariates.
3. Additional non-sampling error not explained by the fixed effects is usually spatially autocorrelated, and this is represented with a random effects component.

This approach has worked well in many applications, but has major limitations. First, the approach relies on having relevant covariates. These covariates are typically gridded rasters that are derived from satellite imagery (Weiss et al. 2018). Commonly used examples are temperature, greenness, aridity, night time lights, and accessibility. For some diseases, such as malaria, these covariates can explain a large amount of the variation in the data, but for complex, socially determined diseases such as HIV, these covariates can explain very little of the variation in the data. We did simple linear modelling on HIV using the standard suite of geospatial covariates and found very low R^2 values (<0.1) for all countries. Exploiting spatial correlation can increase predictive power and can explain variation. However, for HIV, as shown in the data plots (see Section 4.1), the spatial signal is often very weak, and the raw data are highly heterogeneous. Second, HIV data is extremely “noisy.” This is due in part to the nature of the disease and the difficulties in sampling. If prevalence is low ($<10\%$), a large sample size is needed to generate a reliable estimate of prevalence. In this report, we explored the standard approaches to geospatial mapping, and found that they had a very poor fit with the data.

To circumvent the above issues, aggregation is common. Statistically, aggregation reduces error in the bias-variance trade-off by variance reduction through averaging. This aggregation effect is high intuitive and can be seen as boosting the sample size and thus improving estimates of HIV prevalence. From a policy perspective, high resolution maps might be visually appealing, but they do not yield policy relevant information and uncertainty at such a high resolution, particularly for HIV data, is unacceptably high. The main issue with aggregation is that it can be difficult to find relevant covariates that are matched to the same time and spatial resolution, but also share the same sampling design.

In this report, we use a multitask Gaussian process, where we use aggregated spatial coordinates and fit multiple Gaussian processes to HIV prevalence and other indicators, such as women who reported condom use at last high-risk sex with a non-cohabiting, non-marital partner or the number of partners in lifetime for women. However, instead of these Gaussian processes being independent, we also model the cross-correlations between all indicators. This methodology therefore allows for all indicators to inform each

other and increase the predictive power of the model. In addition to prediction, the joint multitask framework also allows for a unified treatment of uncertainty that provides robust uncertainty intervals.

Alternative formulations could be to fit separate Gaussian processes for non-HIV indicators and then use these indicators in a HIV Gaussian process model. However, our attempts at this resulted in severe overfitting of noise in the data. In addition, propagation of uncertainty was impossible.

We note that our approach is most similar to that by Gutreuter and colleagues (Gutreuter et al. 2019) who employ a small area Fay-Herriot estimator to shrink data estimates of HIV prevalence. While our approach also performs shrinkage (albeit using a different functional norm), we can predict to regions which have no data, and can leverage other survey data. We believe our approach shares the strengths of the small area estimation (SAE) approach but provides much additional utility.

3.2 Modelling Limitations

All modelling choices have costs and benefits. The limitations of our framework are, (i) not using the pixel level data but aggregated data, (ii), using a Gaussian likelihood rather than a Binomial likelihood, and (iii) using a continuous Gaussian process rather than a discrete Gaussian Markov random field (GMRF). We justify these limitations as follows: (i) As described in Section 3.1, the noise and sample sizes of the data, together with HIV heterogeneity, mean that pixel level modelling is unadvisable. Errors are unacceptably high, and we do not believe a mathematical method will overcome the limitations inherent in the data. (ii) The correct likelihood for the survey data is the Binomial likelihood, although to allow for feasible accurate computation, we had to use a Gaussian likelihood. However, given the large sample sizes from aggregation, and guarantees from the central limit theorem, a Gaussian likelihood is a highly suitable choice (Stanton and Diggle 2013). (iii) An exponential kernel function was used, which is a close approximation for discrete condition autoregressive (CAR) models (Lindgren, Rue, and Lindstrom 2011).

3.3 Modelling Framework

Given a set of n aggregated spatial coordinates from a DHS survey X , i.e $(lat_i, long_i)$ for $i = [1, \dots, n]$ there are a set of responses. These $m = 8$ responses are described in the ‘‘Data Preparation’’ section. Briefly they are (A) currently married or living in a union, (B) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (C) ever had sexual intercourse between age 15-24, (D) number of partners in lifetime, (E) partners live elsewhere, (F) reporting an STI, (G) reporting an STI, genital discharge, or a sore or ulcer and (H) HIV positivity. Therefore, the m tasks create a response vector $y = [y_{11}, \dots, y_{nm}]$. Define a particular point as k and a particular indicator (or task) as l . The Gaussian process over latent functions, f_l (assuming zero or constant mean), is formed from an inner product space $\langle f_l(x) f_k(x') \rangle = K_{lk}^f k^x(x, x')$ and then from our Gaussian likelihood $y_{il} \sim Normal(f_l(x_i), \sigma_l^2)$. Here K is a positive semi-definite matrix the captures inter-task similarities. k is the standard Gaussian process covariance function that is defined over inputs. σ is the noise, which is different for each indicator. When performing inference and predicting to districts that have no data, for a new coordinate, x^* , predictions are made by $f_l(x^*) = (k_l^f \otimes k_*^x)^T \Sigma^{-1} y$, where $\Sigma = K^f \otimes K^x + D \otimes I$. Here D is a diagonal noise matrix and \otimes is the Kronecker product. This Kronecker helps keeps the model computational tractable. The objective function used for learning all parameters was available due to the Gaussian likelihood.

$$L = -\frac{N}{2} \log |K^f| - \frac{M}{2} |K^x| - \frac{1}{2} \text{tr} \left[(K^f)^{-1} F^T (K^x)^{-1} F \right] - \frac{N}{2} \sum_{l=1}^M \log \sigma_l^2 - \frac{1}{2} \text{tr} [(Y - F) D^{-1} (Y - F)^T] - \frac{MN}{2} \log 2\pi$$

This equation is a marginal likelihood, which means that we integrate over all possible functions in the multitask model. This prevents overfitting and yields reliable estimates of uncertainty. Fitting was performed in Gpytorch (Gardner et al. 2018) using exact GP computation via conjugate gradients. Optimization was performed with adaptive gradient descent Adam optimiser (Kingma and Ba 2014). Prediction and plotting were done using the “ggplot2” package in the R software (R Core Team 2019).

4 RESULTS

4.1 Correlation Plots

To justify and present the utility of the multitask approach, we show correlation plots of the indicators for all countries. In all examples, substantial correlation exists to be exploited by the multitask GP.

Figure 9 Correlation plot for Burundi 2016. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.

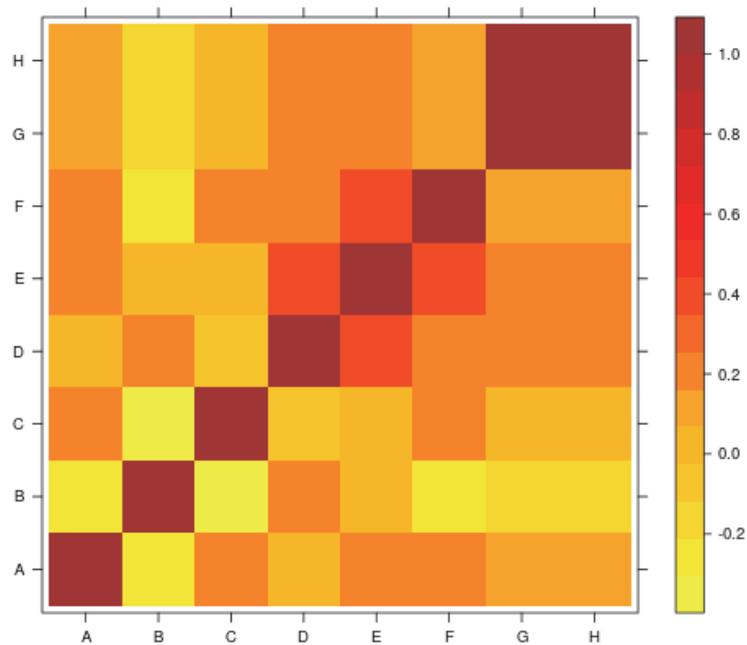


Figure 10 Correlation plot for DRC 2013. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.

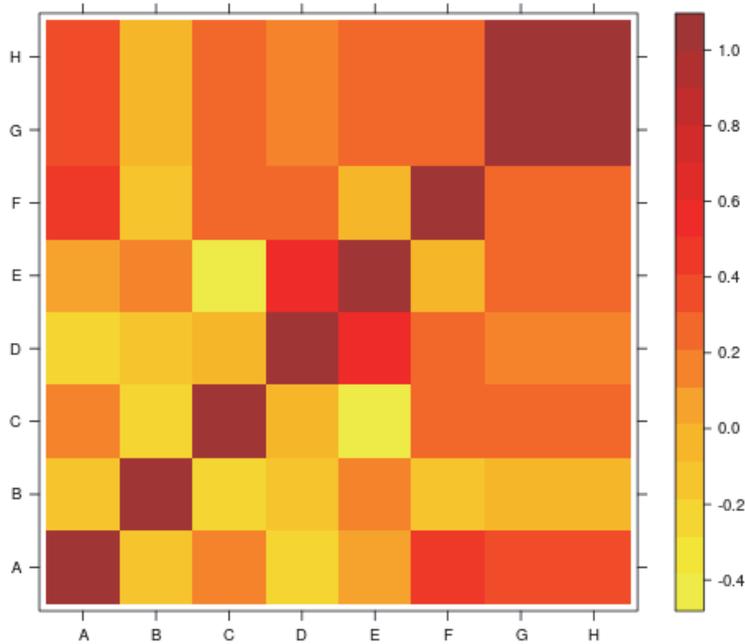


Figure 11 Correlation plot for Ethiopia 2011. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.

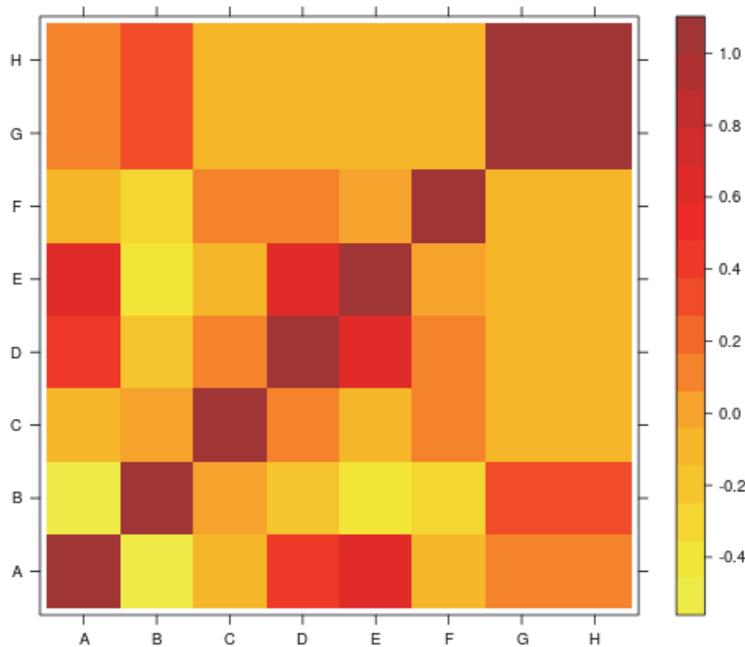


Figure 12 Correlation plot for Ethiopia 2016. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.

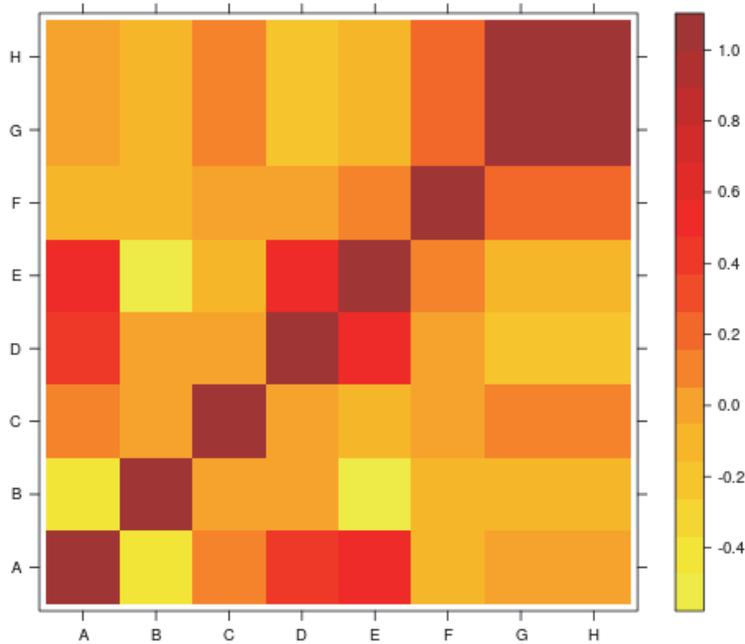


Figure 13 Correlation plot for Ghana 2014. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.

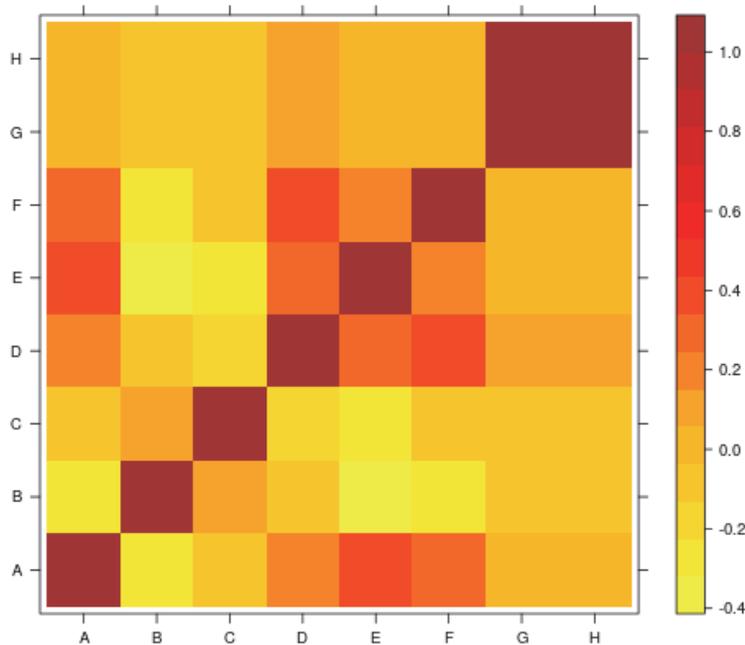


Figure 14 Correlation plot for Mali 2012. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.

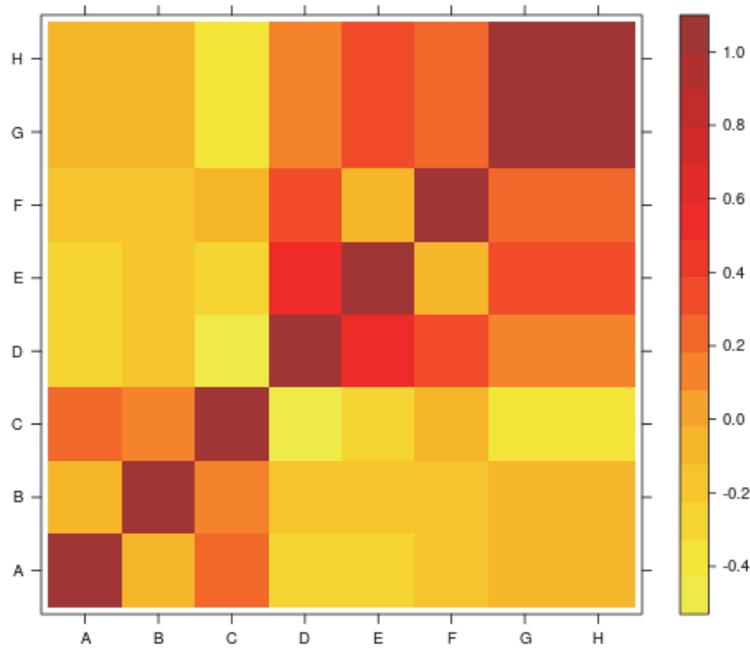


Figure 15 Correlation plot for Rwanda 2015. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.

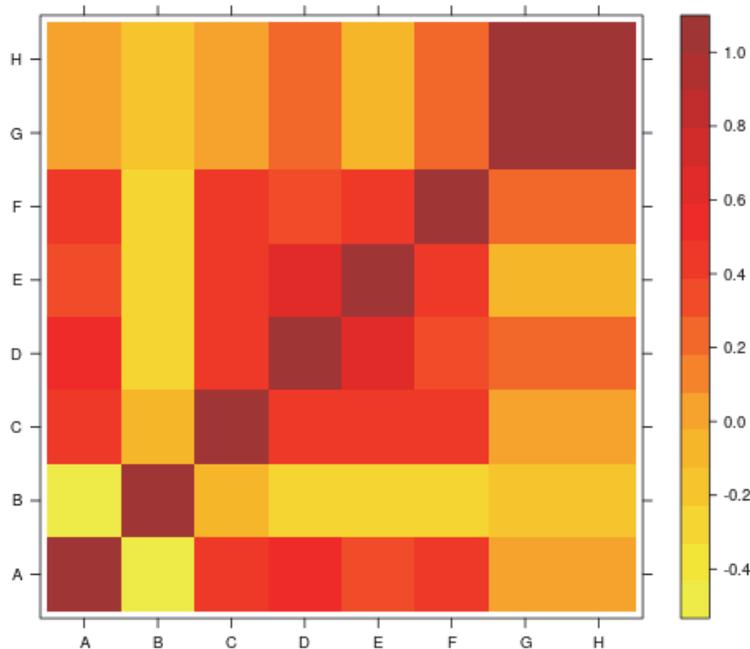
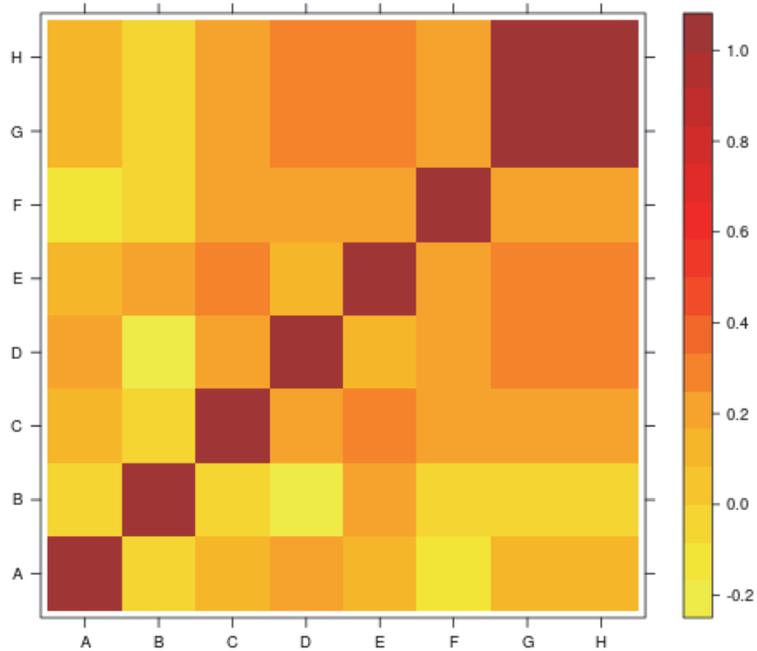


Figure 16 Correlation plot for South Africa 2016. (A) HIV positivity, (B) currently married or living in a union, (C) reported condom use at last high-risk sex with a non-cohabiting, non-marital partner, (D) ever had sexual intercourse between age 15-24, (E) number of partners in lifetime, (F) partners live elsewhere, (G) reporting an STI, (H) reporting an STI, genital discharge, or a sore or ulcer.



4.2 HIV Prevalence Maps and Confidence Interval Plots

Figures 17 to 24 show the ADMIN 2 level estimates that highlight areas with high and low HIV prevalence and the corresponding 95% credible interval for each country that we modeled. The maps are generated with our novel methodology. The multitask GP performs a marginalization task. This marginalization essentially looks at every single possible function from an infinite collection and assigns a weight of how appropriately it fits the data. In this way, we are guarded against overfitting. Another nice property is that we obtain Bayesian credible intervals of uncertainty that grow where there is heterogeneity in data (like a wide spectrum of HIV prevalence) or when there are no data to support prediction. In the maps below, we summarize the mean and the 95% credible interval, which shows the interval where we are sure the HIV prevalence lies between. To interpret the mean and confidence intervals, it is helpful to think that if you went to a district and sampled individuals thousands of times, the average would be the mean, and the confidence interval would be the 2.5% and 97.5% quartiles.

Figure 17 Burundi 2016. Prevalence map (left) and uncertainty (right).

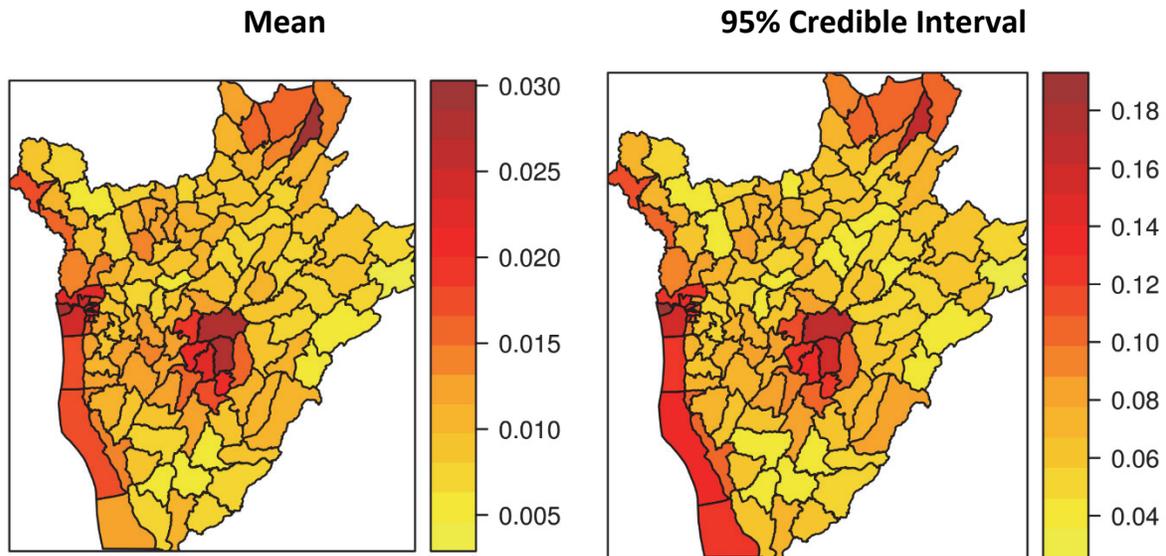


Figure 18 DRC 2016. Prevalence map (left) and uncertainty (right).

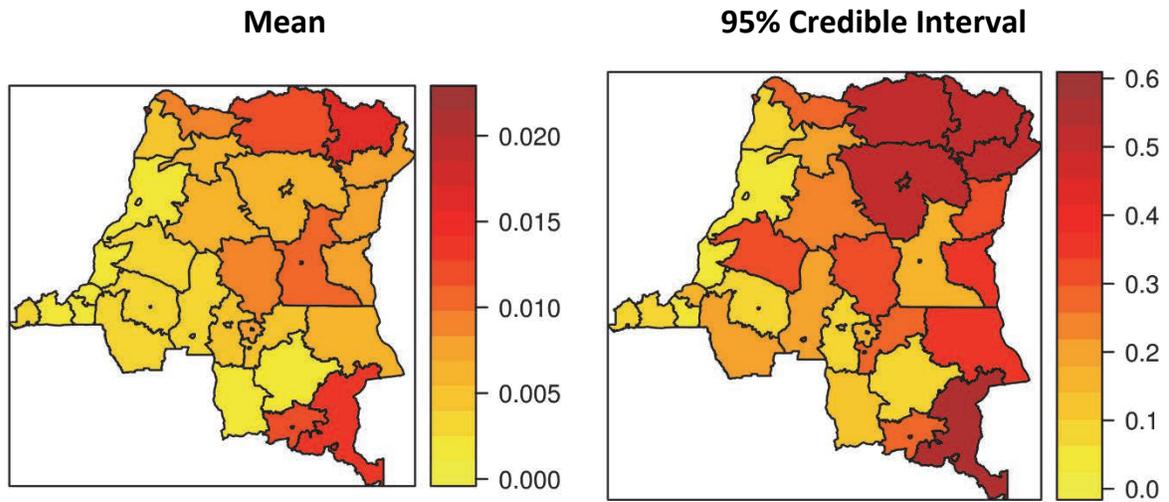


Figure 19 Ethiopia 2011. Prevalence map (left) and uncertainty (right).

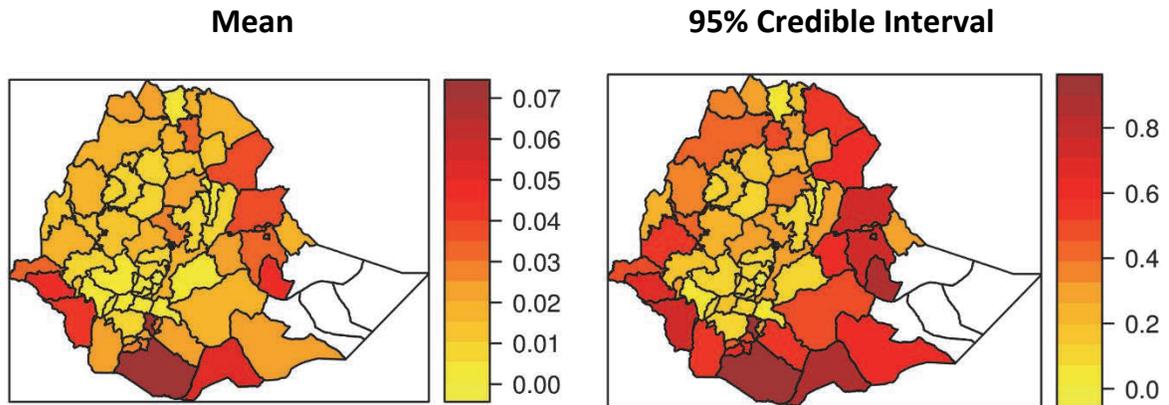


Figure 20 Ethiopia 2016. Prevalence map (left) and uncertainty (right).

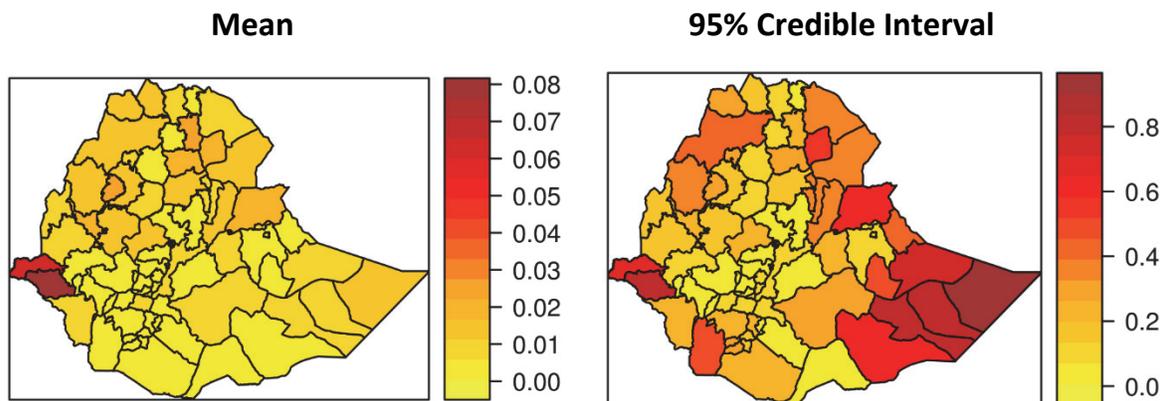


Figure 21 Ghana 2014. Prevalence map (left) and uncertainty (right).

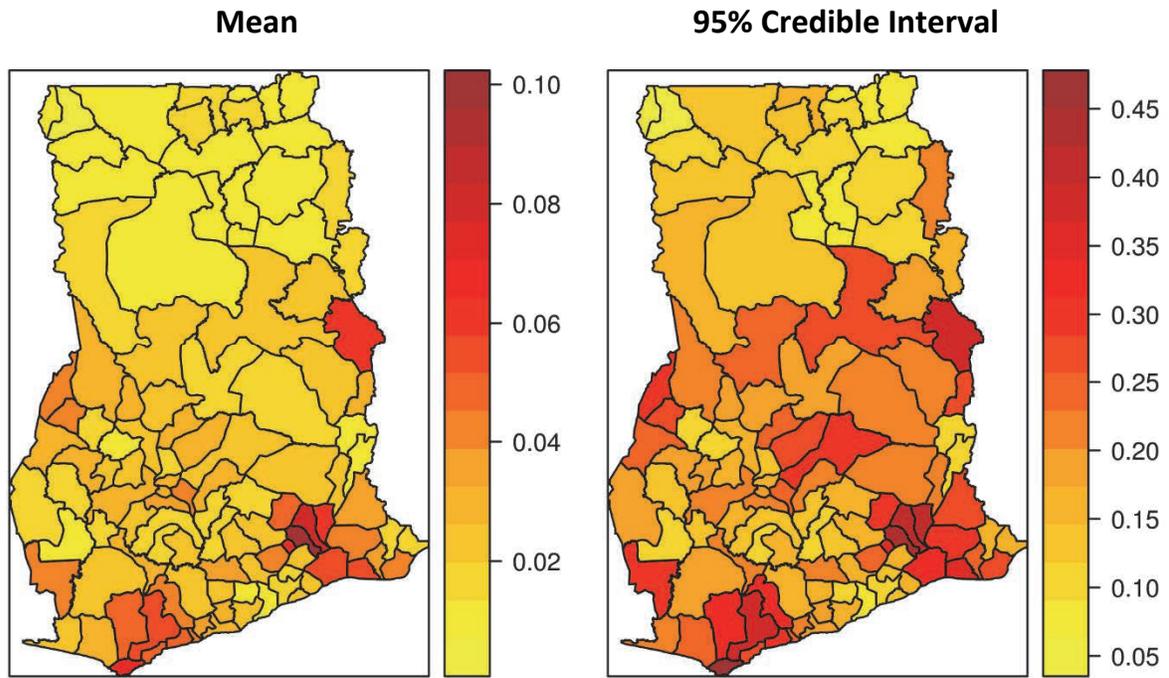


Figure 22 Mali 2012. Prevalence map (left) and uncertainty (right).

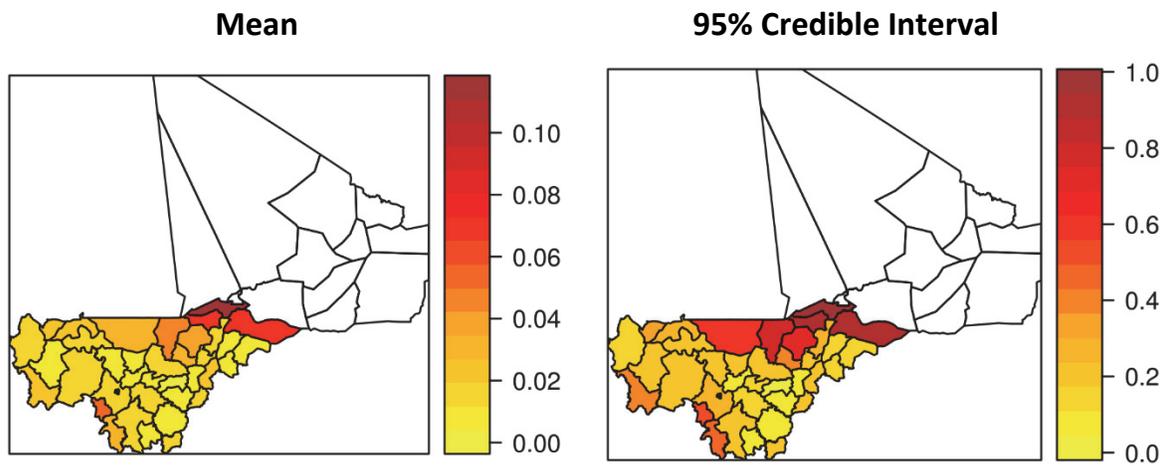


Figure 23 Rwanda 2015. Prevalence map (left) and uncertainty (right).

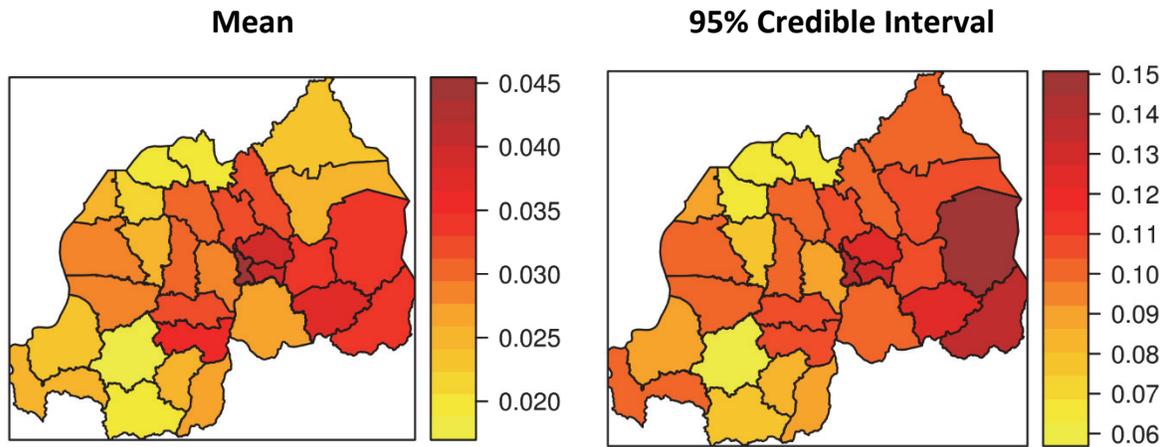
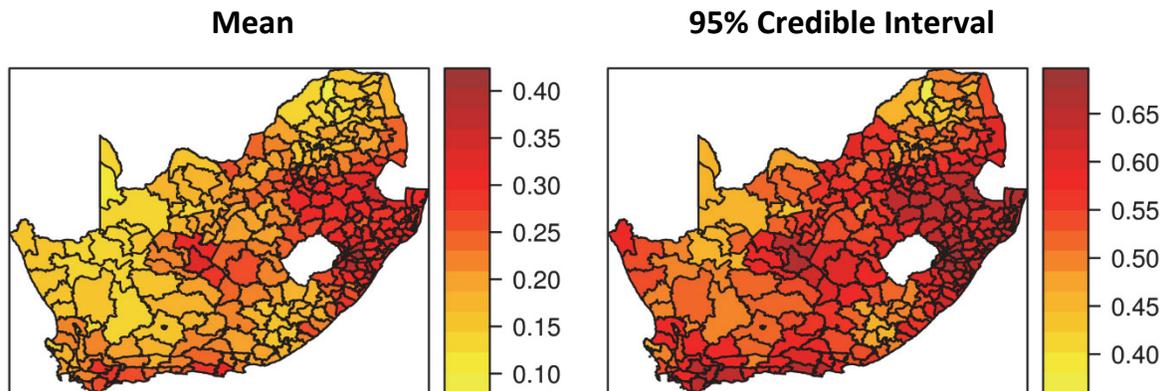


Figure 24 South Africa 2016. Prevalence map (left) and uncertainty (right).



5 CONCLUSION

In this analysis, we have implemented a new approach for HIV prevalence mapping at the ADMIN 2 level. For the first time, this approach uses indicators actually relevant to predicting HIV prevalence. We utilized a Bayesian framework and therefore have a robust treatment of uncertainty. We see many applications of this approach to other diseases in the future.

The generated estimates of HIV prevalence provide essential information about those areas that are potentially in need of more attention. By using the estimates, HIV interventions and programs can be implemented and directed at much smaller spatial scales such as described in our analysis, and can enable better programmatic decisions and targeting of primary prevention strategies.

REFERENCES

- Bor, J., A. J. Herbst, M. L. Newell, and T. Barnighausen. 2013. "Increases in Adult Life Expectancy in Rural South Africa: Valuing the Scale-up of HIV Treatment." *Science* 339 (6122): 961-965. <https://science.sciencemag.org/content/339/6122/961.full>.
- Coburn, B. J., J. T. Okano, and S. Blower. 2017. "Using Geospatial Mapping to Design HIV Elimination Strategies for sub-Saharan Africa." *Science Translational Medicine* 9 (383). <https://doi.org/10.1126/scitransmed.aag0019>.
- Cressie, N. and C. K. Wikle. 2011. *Statistics for Spatio-temporal Data*. Hoboken, NJ, USA: Wiley. <https://doi.org/10.1111/j.1538-4632.2012.00859.x>.
- Cuadros, D. F. and L. J. Abu-Raddad. 2014. "Spatial Variability in HIV Prevalence Declines in Several Countries in sub-Saharan Africa." *Health & Place* 28: 45-49. <https://doi.org/10.1016/j.healthplace.2014.03.007>.
- Cuadros, D. F., B. Sartorius, C. Hall, A. Akullian, T. Bärnighausen, and F. Tanser. 2018. "Capturing the Spatial Variability of HIV Epidemics in South Africa and Tanzania using Routine Healthcare Facility Data." *International Journal of Health Geographics* 17 (1): 27. <https://doi.org/10.1186/s12942-018-0146-8>.
- Dwyer-Lindgren, L., M. A. Cork, A. Sligar, K. M. Steuben, K. F. Wilson, N. R. Provost, B. K. Mayala, et al. 2019. "Mapping HIV Prevalence in sub-Saharan Africa between 2000 and 2017." *Nature* 570 (7760): 189-193. <https://doi.org/10.1038/s41586-019-1200-9>.
- Gardner, J., G. Pleiss, K. Q. Weinberger, D. Bindel, and A. G. Wilson. 2018. "Gpytorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration." Poster presented at the 32nd *Neural Information Processing Systems Conference, Montreal, Canada, December 2-8, 2018*. <http://papers.nips.cc/paper/7985-gpytorch-blackbox-matrix-matrix-gaussian-process-inference-with-gpu-acceleration.pdf>.
- Gething, P. W., A. P. Patil, D. L. Smith, C. A. Guerra, I. R. F. Elyazar, G. L. Johnston, A. J. Tatem, and S. I. Hay. 2011. "A New World Malaria Map: Plasmodium Falciparum Endemicity in 2010." *Malaria Journal* 10 (378). <https://doi.org/10.1186/1475-2875-10-378>.
- Gething, P. W., A. J. Tatem, T. Bird, and C. Burgert. 2015. *Creating Spatial Interpolation Surfaces with DHS Data*. DHS Spatial Analysis Reports 11. Rockville, MD, USA: ICF International. <https://dhsprogram.com/pubs/pdf/SAR11/SAR11.pdf>.
- Gosoni, L., A. Msengwa, C. Lengeler, and P. Vounatsou. 2012. "Spatially Explicit Burden Estimates of Malaria in Tanzania: Bayesian Geostatistical Modeling of the Malaria Indicator Survey Data." *PLOS ONE* 7 (5): e23966. <https://doi.org/10.1371/journal.pone.0023966>.
- Gosoni, L., P. Vounatsou, N. Sogoba, and T. Smith. 2006. "Bayesian Modelling of Geostatistical Malaria Risk Data." *Geospatial Health* 1 (1): 127-139. <https://doi.org/10.4081/gh.2006.287>.

- Gosoni, L., A. M. Veta, and P. Vounatsou. 2011. "Bayesian Geostatistical Modeling of Malaria Indicator Survey Data in Angola." *PLOS ONE* 5 (3): e9322. <https://doi.org/10.1371/journal.pone.0009322>.
- Guttreuter, S., E. Igumbor, N. Wabiri, M. Desai, and L. Durand. 2019. "Improving Estimates of District HIV Prevalence and Burden in South Africa using Small Area Estimation Techniques." *PLOS ONE* 14 (2): e0212445. <https://doi.org/10.1371/journal.pone.0212445>.
- Hallett, T. B., S-J. Anderson, C. A. Asante, N. Bartlett, V. Bendaud, S. Bhatt, C. R. Burgert, D. F. Cuadros, J. Dzangare, D. Fecht, et al. 2016. "Evaluation of Geospatial Methods to Generate Subnational HIV Prevalence Estimates for Local Level Planning." *AIDS* 30 (9): 1467 - 1474. <https://insights.ovid.com/article/00002030-201606010-00017>.
- Hay, S. I., C. A. Guerra, P. W. Gething, A. P. Patil, A. J. Tatem, A. M. Noor, C. W. Kabaria, et al. 2009. "A World Malaria Map: Plasmodium Falciparum Endemicity in 2007." *PLOS MED* 6 (3): e1000048. <https://doi.org/10.1371/journal.pmed.1000048>.
- Kandala, N. B., E. K. Campbell, S. D. Rakgoasi, B. C. Madi-Segwagwe, and T. T. Fako. 2012. "The Geography of HIV/AIDS Prevalence Rates in Botswana." *HIV/AIDS* 4: 95-102. <https://doi.org/10.2147/HIV.S30537>.
- Kazembe, L., I. Kleinschmidt, T. Holtz, and B. Sharp. 2006. "Spatial Analysis and Mapping of Malaria Risk in Malawi using Point-referenced Prevalence of Infection Data." *International Journal of Health Geographics* 5 (1): 41. <https://doi.org/10.1186/1476-072X-5-41>.
- Kingma, D. P. and J. Ba. 2015. "Adam: A Method for Stochastic Optimization." Paper presented at the *3rd International Conference for Learning Representations, San Diego, CA, USA, May 7-9, 2015*. <https://arxiv.org/pdf/1412.6980.pdf>.
- Kleinschmidt, I., A. Pettifor, N. Morris, C. MacPhail, and H. Rees. 2007. "Geographic Distribution of Human Immunodeficiency Virus in South Africa." *The American Journal of Tropical Medicine and Hygiene* 77 (6): 1163-1169. <https://doi.org/10.4269/ajtmh.2007.77.1163>.
- Larmarange, J., R. Vallo, S. Yaro, P. Msellati, and N. Meda. 2011. "Methods for Mapping Regional Trends of HIV Prevalence from Demographic and Health Surveys (DHS)." *Cybergeo: European Journal of Geography*. <https://doi.org/10.4000/cybergeo.24606>.
- Larmarange, J., and V. Bendaud. 2014. "HIV Estimates at Second Subnational Level from National Population-based Surveys." *AIDS* 28 (Suppl 4): S469-76. <https://insights.ovid.com/article/00002030-201411004-00008>.
- Li, Z., Y. Hsiao, J. Godwin, B. D. Martin, J. Wakefield, S. J. Clark, and with support from the United Nations Inter-agency Group for Child Mortality Estimation and its technical advisory group. 2019. "Changes in the Spatial Distribution of the Under-five Mortality Rate: Small-area Analysis of 122 DHS Surveys in 262 Subregions of 35 Countries in Africa." *PLOS ONE* 14 (1): e0210645-e0210645. <https://doi.org/10.1371/journal.pone.0210645>.

- Lindgren, F., H. Rue, and J. Lindstrom. 2011. "An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach." *Journal of Royal Statistical Society Series B* 73 (4): 423-498. <https://doi.org/10.1111/j.1467-9868.2011.00777.x>.
- Mayala, B. K., T. Dontamsetti, T. D. Fish, and T. N. Croft. 2019. *Interpolation of DHS Survey Data at Subnational Administrative Level 2*. DHS Spatial Analysis Reports 17. Rockville, MD, USA: ICF International. <https://dhsprogram.com/publications/publication-sar17-spatial-analysis-reports.cfm>.
- Meyer-Rath, G., J. B. McGillen, D. F. Cuadros, T. B. Hallett, S. Bhatt, N. Wabiri, F. Tanser, and T. Rehle. 2018. "Targeting the Right Interventions to the Right People and Places: The Role of Geospatial Analysis in HIV Program Planning." *AIDS* 32 (8): 957-963. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5965918/>.
- Nakagawa, F., M. May, and A. Phillips. 2013. "Life Expectancy Living with HIV: Recent Estimates and Future Implications." *Current Opinion in Infectious Diseases* 26 (1): 17-25. <https://doi.org/10.1097/QCO.0b013e32835ba6b1>.
- Ngesa, O., H. Mwambi, and T. Achia. 2014. "Bayesian Spatial Semi-parametric Modeling of HIV Variation in Kenya." *PLOS ONE* 9 (7): e103299. <https://doi.org/10.1371/journal.pone.0103299>.
- Niragire, F., T. N. Achia, A. Lyambabaje, and J. Ntaganira. 2015. "Bayesian Mapping of HIV Infection among Women of Reproductive Age in Rwanda." *PLOS ONE* 10 (3): e0119944. <https://doi.org/10.1371/journal.pone.0119944>.
- Noor, A. M., P. W. Gething, V. A. Alegana, A. P. Patil, S. I. Hay, E. Muchiri, E. Juma, and R. W. Snow. 2009. "The Risks of Malaria Infection in Kenya in 2009." *BMC Infectious Diseases* 9 (180). <https://doi.org/10.1186/1471-2334-9-180>.
- Okango, E., H. Mwambi, and O. Ngesa. 2016. "Spatial Modeling of HIV and HSV-2 among Women in Kenya with Spatially varying Coefficients." *BMC Public Health* 16 (1): 355. <https://doi.org/10.1186/1471-2334-9-180>.
- R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria.: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Raso, G., N. Schur, J. Utzinger, B. Koudou, E. Tchicaya, F. Rohner, E. N'Goran, et al. 2012. "Mapping Malaria Risk among Children in Cote D'Ivoire using Bayesian Geo-statistical Models." *Malaria Journal* 11 (1): 160. <https://doi.org/10.1186/1475-2875-11-160>.
- Riedel, N., P. Vounatsou, J. M. Miller, L. Gosoni, E. Chizema-Kawesha, V. Mukonka, and R. W. Steketee. 2011. "Geographical Patterns and Predictors of Malaria Risk in Zambia: Bayesian Geostatistical Modelling of the 2006 Zambia National Malaria Indicator Survey (ZMIS)." *Malaria Journal* 9 (37). <https://doi.org/10.1186/1475-2875-9-37>.
- Stanton, M. C., and P. J. Diggle. 2013. "Geostatistical Analysis of Binomial Data: Generalised Linear or Transformed Gaussian Modelling?" *Environmetrics* 24 (3): 158-171. <https://doi.org/10.1002/env.2205>.

- Steele, J. E., P. R. Sundsøy, C. Pezzulo, V.A. Alegana, T.J. Bird, J. Blumenstock, J. Bjelland, et al. 2017. "Mapping Poverty using Mobile Phone and Satellite Data." *Journal of the Royal Society Interface* 14 (127): 20160690. <https://doi.org/10.1098/rsif.2016.0690>.
- Tiwari, N., C. M. S. Adhikari, A. Tewari, and V. Kandpal. 2006. "Investigation of Geo-spatial Hotspots for the Occurrence of Tuberculosis in Almora District, India, using GIS and Spatial Scan Statistic." *International Journal of Health Geographics* 5 (1): 33. <https://doi.org/10.1186/1476-072X-5-33>.
- Utazi, C. E., J. Thorley, V. A. Alegana, J. M. Ferrari, S. Takahashi, C. J. E. Metcalf, J. Lessler, and A. J. Tatem. 2018. "High Resolution Age-structured Mapping of Childhood Vaccination Coverage in Low and Middle income Countries." *Vaccine* 36 (12): 1583-1591. <https://doi.org/10.1016/j.vaccine.2018.02.020>.
- Weiss, D. J., A. Nelson, H. S. Gibson, W. Temperley, S. Peedell, A. Lieber, M. Hancher, et al. 2018. "A Global Map of Travel Time to Cities to Assess Inequalities in Accessibility in 2015." *Nature* 553 (7688): 333-336. <https://doi.org/10.1038/nature25181>.
- World Health Organization (WHO). 2019. *HIV/IDS Factsheets*. Geneva, Switzerland: World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/hiv-aids>.
- Zulu, L. C., E. Kalipeni, and E. Johannes. 2014. "Analyzing Spatial Clustering and the Spatiotemporal Nature and Trends of HIV/AIDS Prevalence using GIS: The Case of Malawi, 1994-2010." *BMC Infectious Diseases* 14 (1): 285. <https://doi.org/10.1186/1471-2334-14-285>.