Supplementary Materials

Towards UNAIDS Fast-Track goals: Targeting priority geographic areas for HIV prevention and care in Zimbabwe

Diego F. Cuadros[1, 2*], Jingjing Li[1], Zindoga Mukandavire[3], Godfrey N. Musuka[4], Adam J. Branscum[5], Benn Sartorius[6], Owen Mugurungi[7], Frank Tanser[6, 8]

[1]*Deparment of Geography and Geographic Information Science, University of Cincinnati, Cincinnati, USA*

[2]*Health Geography and Disease Modeling Laboratory, University of Cincinnati, Cincinnati, USA*

[3]*Department of Global Health and Development, London School of Hygiene & Tropical Medicine, London, UK*

[4]*ICAP at Columbia University, Harare, Zimbabwe*

[5]*Department of Biostatistics, College of Public Health and Human Sciences, Oregon State University, Corvallis, USA*

[6]*Department of Public Health Medicine, School of Nursing and Public Health, University of KwaZulu-Natal, Durban, South Africa*

[7]*Ministry of Health and Child Care, Harare, Zimbabwe*

[8]*Africa Health Research Institute, University of KwaZulu-Natal, Durban, South Africa*

*To whom correspondence should be addressed. E-mail: diego.cuadros@uc.edu

Supplementary methods

Study area and data sources

The study area was Zimbabwe, a landlocked country in southern Africa. Zimbabwe has the sixth highest HIV prevalence in sub-Saharan Africa (SSA), with approximately 1.2 million persons aged 15-64 years old living with HIV in 2016 [1]. The main source of data for this study was the Zimbabwe Demographic and Health Survey (ZDHS) conducted in 2015 [2]. Subjects were enrolled in the ZDHS via a two-stage sampling procedure to select households. A total of 400 ZDHS sample locations were selected. The global positioning system was used to identify and record the geographical coordinates of each ZDHS sample location. The study population was limited to adults aged 15 to 49 years, resulting in a total of 9,054 women and 7,082 men included in the study. Anonymous HIV testing was performed with the informed consent of all sampled individuals. Further details related to the ZDHS methodology, study design, and data can be found elsewhere [2-5].

Population density throughout Zimbabwe was obtained from the WorldPop database (Figure S1). This database was used to extract estimates of the 2014 population density in Zimbabwe [6]. Further details of the methodology used by WorldPop are described elsewhere [6, 7]. Population density maps for males and females aged 15- 49 years were generated using estimates for 2014 in Zimbabwe and population age-distribution projections from the Zimbabwe Population Projections Thematic Report [8].

Estimates for additional spatial variables, including the percent of people living with HIV (PLHIV) who were diagnosed, the percent of people diagnosed with HIV who were treated by ART, and the percent on ART who were virally suppressed, were obtained at provincial scale from the 2016 Zimbabwe Population-Based HIV Impact Assessment (ZIMPHIA) report [1]. ZIMPHIA 2016 was a nationally representative cross-sectional population-based survey of households across Zimbabwe conducted between October 2015 and August 2016. Similar to ZDHS, ZIMPHIA used a two-stage stratified cluster sample design. The number of selected households was calculated to provide a representative provincial estimate of HIV incidence and viral suppression among HIV-positive individuals, as well as other HIV-related epidemiological measures among persons aged 15 to 64 years. A total of 11,098 males and 14,033 females aged 15 to 64 years were eligible to participate in

ZIMPHIA. Questionnaire and field laboratory data were collected among participants. The adult interview included modules on demographics, sexual and reproductive health, sexual activity, and the HIV continuum of care, among other health issues. Data were aggregated, and results were reported at province level. Further details related to the ZIMPHIA methodology, study design, and data can be found elsewhere [1].

The estimated number of new HIV infections and the average number of clients per ART treatment site at the province level were obtained from the report on Smart Investment to End HIV AIDS in Zimbabwe based on Hotspot Analysis [9]. The hotspot analysis was a project aimed to review and analyze populations at greatest risk of HIV infection. An analysis of peer reviewed publications, surveys and published articles was conducted in 2015 to explore HIV hotspots through the identification of HIV drivers in Zimbabwe as well as interventions that have been put in place to curb the epidemic. Further details related to the methodology, study design, and data can be found elsewhere [9].

Definition of socio-economic and behavioral variables

To assess the association between socio-economic and behavioral factors and the prevalence of HIV infection, we selected a variety of demographic, socio-economic, and behavioral attributes from the ZDHS dataset and other sources. Fifteen variables measured for men and 14 for women included: wealth index, number of children under five years old in household (women only), male circumcision (men only), in union status, number of unions, currently reside with husband/wife/partners, number of wives (men only), number of husband's other wives (women only), age at first sex, employment, condom use, number of sexual partners including spouse, ever been tested for HIV, and total number of lifetime sexual partners.

Socio-economic and behavioral variables were quantified as percentages at each ZDHS sample location for males and females separately. For example, for each sampled location, condom use refers to the percentage of individuals who used a condom during the last time having sex, male circumcision refers to the percentage of circumcised males, number of unions is the percentage of individuals with more than one union, number of lifetime sexual partners is the percentage of individuals with less than three lifetime sexual partners, age at first sex corresponds to the percentage

of females who experienced sexual activity at age < 20 years old, and level of education is the percentage of individuals with secondary or higher education. Wealth index is an ordinal variable that describes standard of living as determined by material possessions. This index was used to calculate a poverty variable as the percent of poorest and poorer people.

Definition of environmental variables

Environmental variables included the normalized difference vegetation index (NDVI), distance to primary roads, distance to main cities, and population density (Supplementary Figure S1). NDVI serves as a measure of the degree of urbanization at a sampled location, and it has been shown to be associated with the occurrence of coinfection with a parasitic disease (malaria) that could lead to increased HIV transmission efficiency in SSA [10-12]. NDVI (raster) data were obtained from NASA's Earth Observatory Group [13]. Primary roads and main cities (vector) data were downloaded from DIVA-GIS [14]. A raster layer of distance to primary roads was calculated as the Euclidean distance for each pixel to the nearest main road. Similarly, a raster layer of distance to main cities was measured as the Euclidean distance for each pixel to the nearest main city. We further resampled these raster layers to 1-km spatial resolution with the same projection (WGS_1984_UTM_Zone_36S). Values for these variables were extracted at each ZDHS sample location using the corresponding raster layers.

Selection of spatial variables

To map the geographical distribution of HIV prevalence in Zimbabwe, we implemented a method previously developed to describe the spatial structure of HIV prevalence in several SSA countries [15]. First, preliminary bivariate logistic regression analyses were conducted to assess the existence of unadjusted associations between each variable and HIV prevalence. In addition, we explored the spatial correlation of each variable using the Global Moran's I statistic. A screening of spatial variables involved two criteria: (1) the bivariate logistic regression slope coefficient must have *p-value* < 0.1; and (2) the significance test for spatial correlation must have *p-value* < 0.05. Second, multiple logistic regression analysis was used to evaluate associations between socio-economic, behavioral and environmental variables and the prevalence of HIV infection. Starting with a logistic regression model that included all spatial variables that met the screening criteria, backward stepwise

variable selection was applied to determine a final model using a *p-value* threshold of 0.05 for exiting or re-entering the model. Population density was included in the final model regardless of the *p-value*. Logistic regression and spatial correlation analyses were conducted using SAS® version 9.3 [16].

Mapping predictor variables and HIV prevalence

Ordinary kriging was used to predict values of the selected socio-economic and behavioral variables at unmeasured locations by estimating a variogram of weighted averages of the data [17-20], and then generating continuous surface maps of the variables selected in the gender-specific final logistic regression models (Supplementary Figures S2 and S3). These continuous surface maps have the same spatial resolution and projection as the environmental raster layers. Gender-specific HIV prevalence maps were then generated by substituting values from all continuous surface maps into the gender-specific multivariable logistic regression model using Map Algebra and the logistic equation,

$$HIV\ Prevalence = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}}{1+e^{\beta_0+\beta_1 X_1+\cdots+\beta_k X_k}} \quad (1)$$

where each X_j denotes a predictor variable (e.g., condom use percentage) and β_j is the corresponding regression coefficient. This method, which was used to generate a continuous map of HIV prevalence in raster format with 1-kilometer grid resolution for each gender, was implemented in ArcGIS version 10.3 [21].

Geographic dispersion of HIV-infected males and females

Estimates of population densities for Zimbabwe in 2014 from the WorldPop dataset were obtained in the form of a raster image with a resolution of 100 m x 100 m. For generalization and to reduce the computational burden, we reduced the resolution of the raster file to 5 km x 5 km to estimate the number of HIV infected individuals per 5 km^2. The geographical distributions of the estimated population densities of HIV-infected females and males were generated by multiplying values from the gender-specific population density maps with values from the corresponding gender-specific HIV prevalence maps for females and males that were obtained using equation 1.

High burden areas where both HIV prevalence and the number of HIV-infected individuals were high (spatially clustered) were identified using an analysis of bivariate local indicators of spatial autocorrelation (LISA) that was implemented in the GeoDa platform [22]. Bivariate LISA analysis provides a local assessment of spatial correlation by determining whether a value in a given location is more similar to that of the average of its neighbors, assuming spatial randomness. The bivariate LISA statistic identifies significant spatial clustering (autocorrelation) by measuring the degree of linear association between a variable at a given location (e.g., HIV prevalence) and the average of another variable (e.g., density of HIV-infected individuals) at neighboring (spatially lagged) locations [23]. Gender-specific maps were generated to illustrate the locations where HIV-infected females and males were clustered in areas with high HIV prevalence.

Spatial distribution of HIV treatment and care

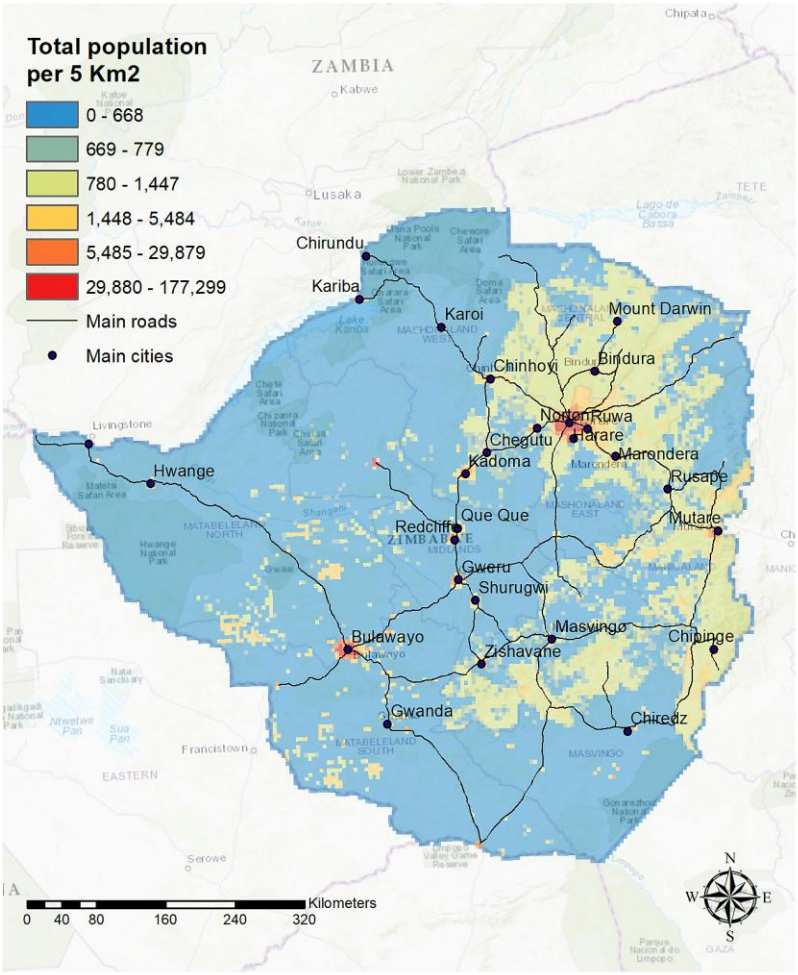
The spatial distributions of HIV incidence in 2013 and average number of clients per ART treatment site were estimated using data from the report on Smart Investment to End HIV AIDS in Zimbabwe Based on the Hotspot Analysis [24]. Data at the provincial levels on the percent of PLHIV who have been diagnosed, the percent diagnosed with HIV who were receiving ART, and the percent of those receiving ART who were virally suppressed were generated by combining estimates from the 2016 ZIMPHIA report [1] with the geographical densities of HIV-infected males and females obtained in this study to determine the spatial distributions of the number of males and females who were lacking HIV treatment and care in Zimbabwe.

Specifically, the percent of PLHIV who were undiagnosed was multiplied by population density estimates of HIV-infected females and males to calculate the number of undiagnosed PLHIV. The percent of PLHIV who were diagnosed but not receiving ART was calculated at the province level as the complement of the percent of those diagnosed and taking ART. This proportion was then multiplied by values in the gender-specific population density maps of diagnosed PLHIV to calculate the number of HIV-infected males and females who were not receiving ART treatment. The percent of individuals on ART but not virally suppressed was calculated at the province level as the complement of the percent of those on ART and virally suppressed. This proportion was multiplied by values in the gender-specific population density maps of the number of males and females on

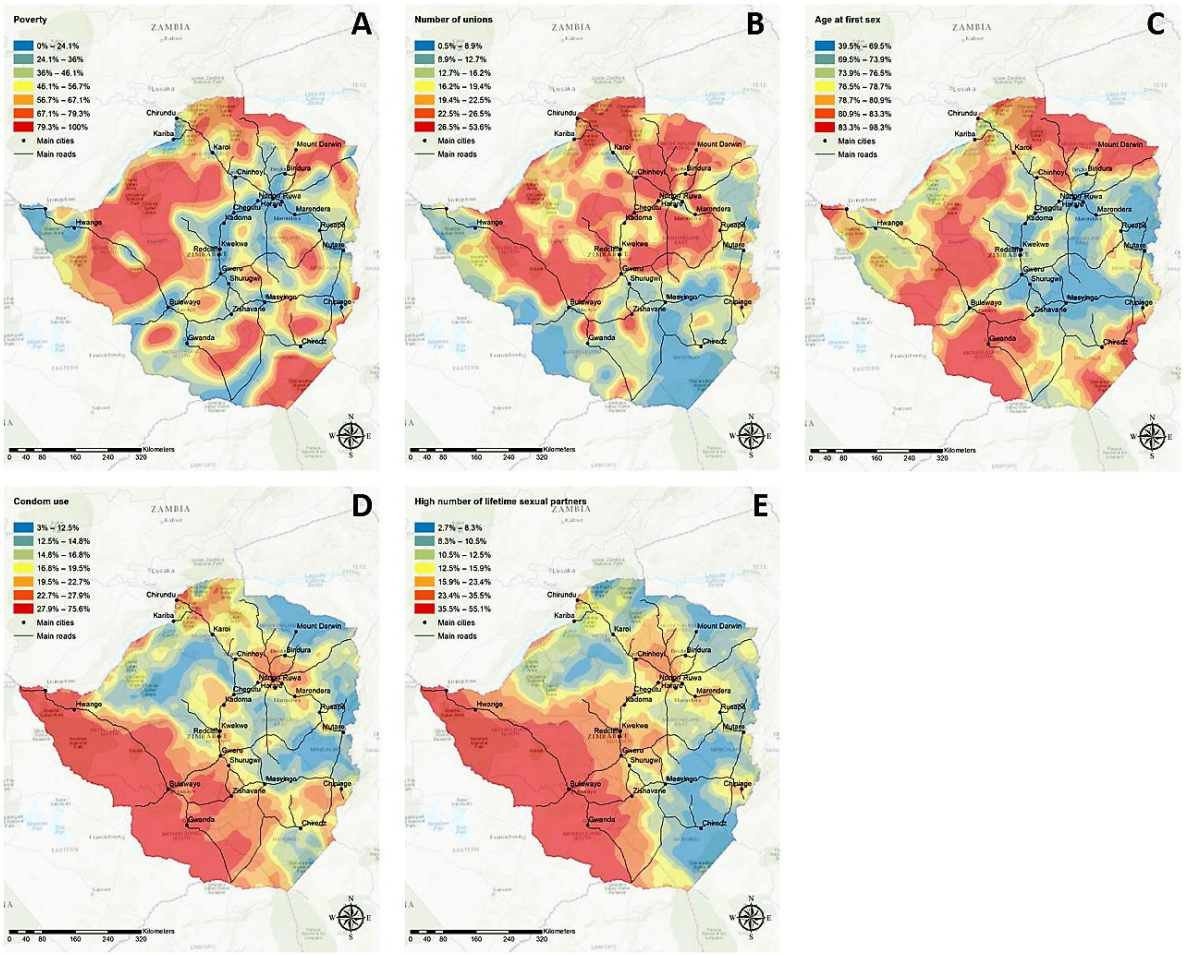
ART to calculate the number of ART-treated individuals who were not virally suppressed. Finally, the calculated numbers of (i) PLHIV who were undiagnosed, (ii) PLHIV who were diagnosed but not on ART, and (iii) PLHIV who were receiving ART but not virally suppressed were summed to generate gender-specific maps of the geographical distributions of the numbers of males and females aged 15-49 years lacking HIV treatment and care in Zimbabwe in 2015.

Supplementary figures

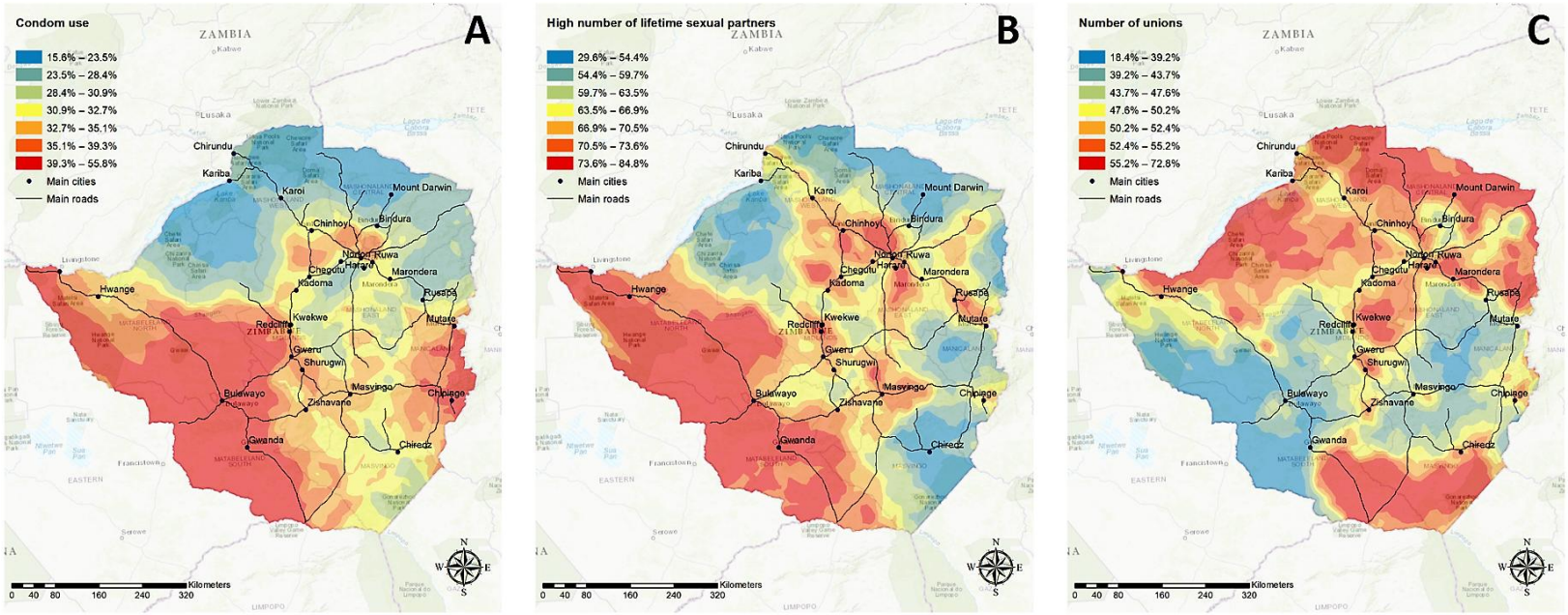
Supplementary Figure S1. Density distribution of the total population in Zimbabwe in 2014. Maps were created using ArcGIS® by ESRI version 10.3 (http://www.esri.com) [21]



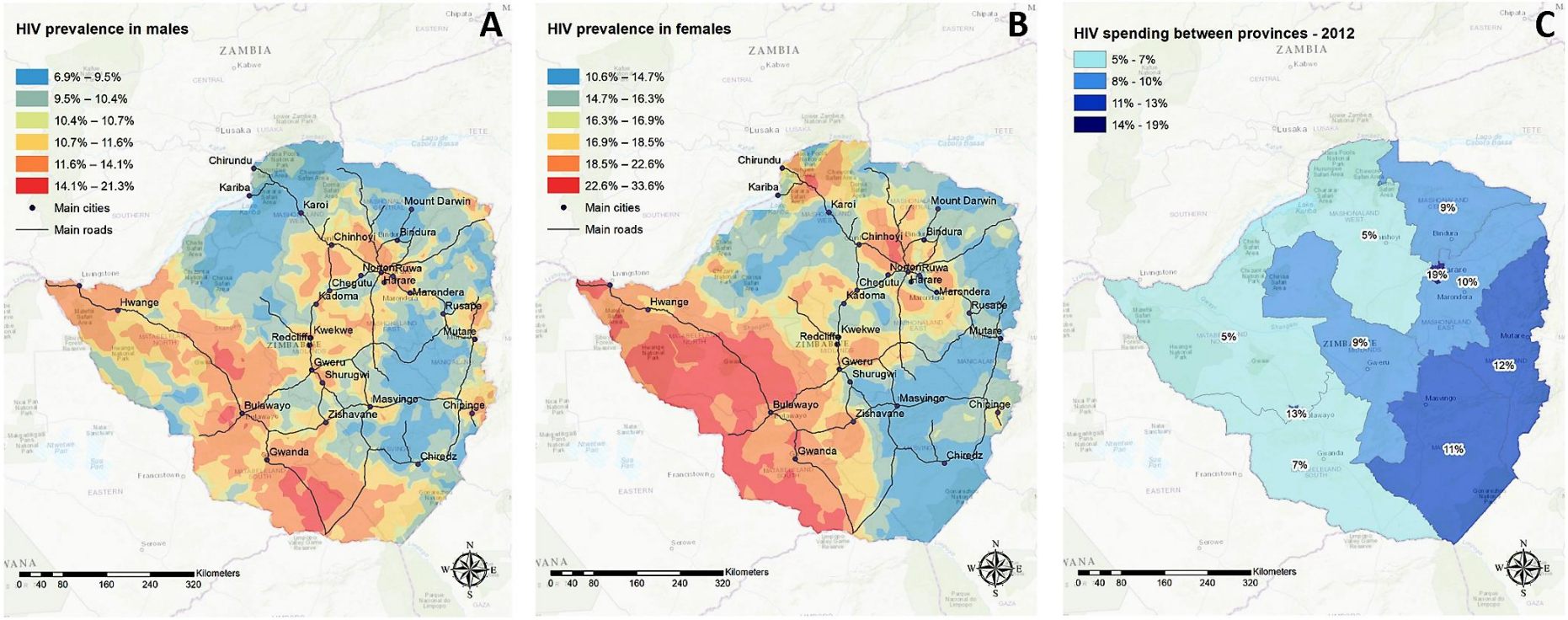
Supplementary Figure S2. Continuous surface maps for the cofactors in the final model for females. Maps were created using ArcGIS® by ESRI version 10.3 (http://www.esri.com) [21]



Supplementary Figure S3. Continuous surface maps for the cofactors in the final model for males. Maps were created using ArcGIS® by ESRI version 10.3 (http://www.esri.com) [21]



Supplementary Figure S4. Continuous surface maps for the HIV prevalence in Zimbabwe for (A) males and (B) females, and HIV spending between provinces in 2012 (C). Maps were created using ArcGIS® by ESRI version 10.3 (http://www.esri.com) [21]



Supplementary tables

Supplementary Table 1. Descriptive statistics of socio-behavioral variables included in the analysis

Gender		Variable	Mean	Standard Deviation
Males	**Socio-behavioral**	*Secondary or higher education (%)*	77.74	21.69
		Poverty (poorer or poorest) (%)	29.66	32
		Modern contraceptive method (%)	56.06	14.31
		Circumcision (%)	16.89	14.92
		Currently in union (%)	49.29	16.91
		Number of unions (more than once) (%)	77.23	16.97
		Currently reside with wife/partner (%)	87.4	14.61
		Number of wives (more than one) (%)	3.75	9.07
		Age at first sex below 20 (%)	57.19	18.47
		Employed (%)	64.51	19.08
		Condom use (%)	32.96	16.52
		Number of sex partners including spouse more than two (%)	2	3.37
		Ever been tested for HIV (%)	66.03	14.5
		Lifetime number of sexual partners more than two (%)	65.27	16.34
	Environmental	*NDVI*	127.14	15.09
		Distance to primary roads (km)	16.26	21.06
		Distance to main cities (km)	44.23	45.28
		Population density (/km^2)	736.24	1217.24
Females	**Socio-behavioral**	*Secondary or higher education (%)*	73.97	21.29
		Poverty (poorer or poorest) (%)	31.89	33.89
		Number of children aged five and under in household (%)	1.15	3.28
		Currently in union (%)	60.92	14.43
		Number of unions (more than once) (%)	16.37	12.16
		Currently reside with husband/partner (%)	74.37	17.26
		Number of women whose husbands have other wives (%)	7.86	9.64

	Age at first sex below 20 (%)	72.77	16.97
	Employed (%)	40.11	16.7
	Condom use (%)	20	13.38
	Number of sex partners including spouse more than two (%)	0.2	0.96
	Ever been tested for HIV (%)	82.71	9.87
	Lifetime number of sexual partners more than two (%)	17.93	14.33
Environmental	*NDVI*	127.14	15.09
	Distance to primary roads (km)	16.26	21.06
	Distance to main cities (km)	44.23	45.28
	Population density (/km^2)	736.24	1217.24

Supplementary Table 2. Bivariate logistic regression and Moran's I analysis for cofactor selection

Gender	Variable	Estimate	P-value	Moran's I	P-value
Males	*Secondary or higher education (%)*	0.000020	0.99	0.27	< 0.001
	Poverty (poorer or poorest) (%)	-0.003907	< 0.001	0.39	< 0.001
	Modern contraceptive method (%)	0.011397	0.00	0.07	< 0.001
	Circumcision (%)	0.000261	0.92	0.23	< 0.001
	Currently in union (%)	0.006444	0.01	0.09	< 0.001
	Number of unions (more than once) (%)	-0.011715	< 0.001	0.00	0.73
	Currently reside with wife/partner (%)	-0.007160	0.01	0.11	< 0.001
	Number of wives more than one (%)	-0.016122	< 0.001	0.02	0.11
	Age at first sex below 20 (%)	0.006106	< 0.001	0.07	< 0.001
	Employed (%)	0.007788	< 0.001	0.10	< 0.001
	Condom use (%)	0.009422	< 0.001	0.16	< 0.001
	Number of sex partners including spouse more than two (%)	0.029610	0.01	0.00	0.83
	Ever been tested for HIV (%)	0.008006	< 0.001	0.06	< 0.001
	Lifetime number of sexual partners more than two (%)	0.014120	< 0.001	0.05	< 0.001
	NDVI	0.001457	0.55	-	-
	Distance to primary roads (km)	-0.000006	0.00	-	-
	Distance to main cities (km)	-0.000001	0.12	-	-
	Population density (/km^2)	0.000038	0.20	-	-
Females	*Secondary or higher education (%)*	-0.000821	0.52	0.26	< 0.001
	Poverty (poorer or poorest) (%)	-0.001312	0.11	0.39	< 0.001
	Number of children aged five and under in household (%)	0.007165	0.39	< 0.001	0.93
	Currently in union (%)	-0.008513	< 0.001	0.17	< 0.001
	Number of unions (more than once) (%)	0.015719	< 0.001	0.04	< 0.001
	Currently reside with husband/partner (%)	-0.001519	0.35	0.22	< 0.001
	Number of women whose husbands have other wives (%)	-0.000297	0.92	0.09	< 0.001
	Age at first sex below 20 (%)	0.007454	< 0.001	0.21	< 0.001
	Employed (%)	0.001683	0.32	0.21	< 0.001
	Condom use (%)	0.020800	< 0.001	0.15	< 0.001

Number of sex partners including spouse more than two (%)	0.058270	0.03	-0.01	0.45
Ever been tested for HIV (%)	0.014860	< 0.001	0.02	0.10
Lifetime number of sexual partners more than two (%)	0.020015	< 0.001	0.31	< 0.001
NDVI	0.000172	0.92	-	-
Distance to primary roads (km)	-0.000001	0.29	-	-
Distance to main cities (km)	0.000001	0.05	-	-
Population density (/km^2)	-0.000013	0.57	-	-

References

1. Ministry of Health and Child Care (MOHCC), Zimbabwe. **Zimbabwe Population-Based HIV Impact Assessment (ZIMPHIA) 2015-16: First Report. Harare, MOHCC** (Available at: https://phia.icap.columbia.edu/wp-content/uploads/2017/11/ZIMPHIA_First_Report_FINAL.pdf). 2017.

2. Zimbabwe National Statistics Agency, ICF International. **Zimbabwe Demographic and Health Survey 2015: Final Report**. In. Rockville, Maryland, USA: Zimbabwe National Statistics Agency (ZIMSTAT) and ICF International (Available at: https://dhsprogram.com/pubs/pdf/FR322/FR322.pdf); 2016.

3. Tanzania Commission for AIDS ZAC, National Bureau of Statistics, Office of Chief Goverment Statistician, ICF international. **Tanzania HIV/AIDS and Malaria Indicator Survey 2011-12**. In: Calverton, MD: ICF International (Available at: https://dhsprogram.com/pubs/pdf/AIS11/AIS11.pdf); 2013.

4. Tanzania Commission for AIDS ZAC, National Bureau of Statistics, Office of Chief Goverment Statistician, Macro International Inc. **Tanzania HIV/AIDS and Malaria Indicator Survey 2007-08**. In: Calverton, MD: Macro International Inc. (Available at: https://dhsprogram.com/pubs/pdf/AIS6/AIS6_05_14_09.pdf); 2008.

5. Tanzania Commission for AIDS NBoS, ORC Macro. **Tanzania HIV/AIDS indicator Survey 2003-04**. In: Calverton, MD: ORC Macro (Available at: https://dhsprogram.com/pubs/pdf/AIS1/AIS1.pdf); 2005.

6. **WorldPop Project (Available at: www.worldpop.org.)**. 2014.

7. Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ. **Population distribution, settlement patterns and accessibility across Africa in 2010**. *PloS one* 2012; 7(2):e31743.

8. The Zimbabwe National Statistics Agency (ZIMSTAT)**. Population Projections Thematic Report** (http://www.zimstat.co.zw/sites/default/files/img/publications/Census/population_projection.pdf). 2015.

9. United Nations World Food Programme, Zimbabwe**. Smart Investment to End HIV AIDS in ZIMBABWE based on Hotspot Analysis** (Available at: http://procurement-notices.undp.org/view_file.cfm?doc_id=114244). 2015.

10. Cuadros DF, Branscum AJ, Crowley PH. **HIV–malaria co-infection: effects of malaria on the prevalence of HIV in East sub-Saharan Africa**. *International journal of epidemiology* 2011; 40(4):931-939.

11. Abu-Raddad LJ, Patnaik P, Kublin JG. **Dual infection with HIV and malaria fuels the spread of both diseases in sub-Saharan Africa**. *Science* 2006; 314(5805):1603-1606.

12. Harms G, Feldmeier H. **HIV infection and tropical parasitic diseases–deleterious interactions in both directions?** *Tropical Medicine & International Health* 2002; 7(6):479-488.

13. NASA. **Earth Observatory. Measuring Vegetation (NDVI & EVI)** (Available at:

http://earthobservatory.nasa.gov/Features/MeasuringVegetation/)

14. DIVA-GIS. **(Available at: http://www.diva-gis.org/gdata)**.

15. Cuadros DF, Li J, Branscum AJ, Akullian A, Jia P, Mziray EN, et al. **Mapping the spatial variability of HIV infection in Sub-Saharan Africa: Effective information for localized HIV prevention and control**. *Scientific reports* 2017; 7(1):9093.

16. **SAS Instituter Inc. SAS. Cary, NC, USA**. 2006.

17. Berke O. **Exploratory disease mapping: kriging the spatial risk function from regional count data**. *International Journal of Health Geographics* 2004; 3(1):18.

18. Carrat F, Valleron A-J. **Epidemiologic mapping using the "kriging" method: application to an influenza-like epidemic in France**. *American journal of epidemiology* 1992; 135(11):1293-1300.

19. Goovaerts P. **Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging**. *International Journal of Health Geographics* 2005; 4(1):31.

20. Naish S, Dale P, Mackenzie JS, McBride J, Mengersen K, Tong S. **Spatial and temporal patterns of locally-acquired dengue transmission in northern Queensland, Australia, 1993–2012**. *PloS one* 2014; 9(4):e92524.

21. ESRI. **ArcGIS 10.x. Redlands, CA, USA: ESRI**. 2004.

22. Anselin L, Syabri I, Kho Y. **GeoDa: an introduction to spatial data analysis**. *Geographical analysis* 2006; 38(1):5-22.

23. Anselin L. **Local indicators of spatial association—LISA**. *Geographical analysis* 1995; 27(2):93-115.

24. United Nations World Food Programme, Zimbabwe. **Smart Investment to End HIV AIDS in ZIMBABWE based on Hotspot Analysis**. 2015.