Geospatial Health 2023; volume 18:1207

Spatial heterogeneity in relationship between district patterns of HIV incidence and covariates in Zimbabwe: a multi-scale geographically weighted regression analysis

Rutendo Birri Makota, Eustasius Musenge

Division of Epidemiology and Biostatistics, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

Abstract

A study was conducted to investigate the district-level patterns of incidence of the human immunodeficiency virus (HIV) in Zimbabwe in the period 2005-2015 and explore variations in the relationship between covariates and HIV incidence across different districts. Demographic health survey data were analysed using hotspot analysis, spatial autocorrelation, and multi-scale geographically weighted regression (MGWR) techniques. The analysis revealed hotspots of the HIV epidemic in the southern and western regions of Zimbabwe in contrast to the eastern and northern regions. Specific districts in Matabeleland South and Matabeleland North provinces showed clusters of HIV incidence in 2005-2006, 2010-2011 and 2015. Variables studied were multiple sex partners and sexually transmitted infections (STI) condom use and being married. Recommendations include implementing targeted HIV prevention programmes in identified hotspots, prioritising interventions addressing the factors mentioned above as well as enhancing access to HIV testing and treatment services in high-risk areas, strengthening surveillance systems, and conducting further research to tailor interventions based on contextual factors. The study also emphasizes the need for regular monitoring and evaluation at the district level to inform effective responses to the HIV epidemic over time. By addressing the unique challenges and risk factors in different districts, significant progress can be made in reducing HIV transmission and improving health outcomes in Zimbabwe. These findings should be valuable for policymakers in resource allocation and designing evidence-based interventions.

Correspondence: Rutendo Birri Makota, Division of Epidemiology and Biostatistics, School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.
E-mail: rutendobbirri@gmail.com

Key words: demographic health survey; HIV incidence; hotspot analysis; multi-scale geographically weighted regression; spatial analysis; Zimbabwe.

Conflict of interest: the authors declare no potential conflict of interest, and all authors confirm accuracy.

Availability of data and materials: all data generated or analyzed during this study are included in this published article.

Received: 26 April 2023.
Accepted: 28 August 2023.

Geospatial Health 2023; 18:1207
doi:10.4081/gh.2023.1207

Introduction

The human immunodeficiency virus (HIV) epidemic continues to pose a significant public health challenge in Sub-Saharan Africa, where nearly 60% of all people living with HIV in the world reside (Awaidy *et al.*, 2023; Jahagirdar *et al.*, 2021). Understanding the spatial heterogeneity in HIV incidence and its relationship with various covariates is crucial for designing targeted interventions and allocating resources effectively. While progress has been made in identifying risk factors associated with HIV transmission at the regional and national levels, the need remains to examine the district-level patterns and their specific determinants.

Numerous studies conducted in Sub-Saharan Africa have explored the relationship between HIV incidence and various covariates, including socio-demographic, behavioural, and structural factors. Bulstra *et al.*, (2020) mapped and characterised areas with high levels of HIV transmission in sub-Saharan Africa and found that the spatial distribution of HIV incidence is associated with factors, such as having more than 10 sexual partners, presence of sexually transmitted infections (STIs), not being circumcised (men only), wealth index, education level and employment status (Bulstra *et al.*, 2020). Other studies conducted in Sub-Saharan Africa examined the associations between HIV incidence and various factors such as age, gender, marital status, condom use and sexual debut age (Coburn *et al.*, 2013; Kharsany & Karim, 2016; Patel *et al.*, 2022). These studies have investigated the impact and highlighted the complex interplay of socio-demographic, behavioural, and economic factors in shaping the spatial patterns of HIV transmission. However, most of the existing research has focused on national-level estimates or aggregated data at larger geographic scales, such as region or province, thereby overlooking the potential spatial heterogeneity that may exist within smaller administrative units. Recent advancements in spatial analysis techniques, coupled with the increased availability of geographically referenced HIV data, enable the exploration of the spatial patterns of HIV incidence at such finer spatial scales (Aturinde *et al.*, 2019; Cuadros *et al.*, 2017; Dwyer-Lindgren *et al.*, 2019; Moturi *et al.*, 2022). These approaches allow the identification of local HIV hotspots as well as coldspots, which may represent areas with lower risk.

Most studies in Zimbabwe have explored the relationship between HIV prevalence and covariates, mainly at national and provincial levels using global regression models. Schaefer *et al.* (2017) explored factors that were likely to influence the equity implications and cost-effectiveness geographically of HIV resource allocation in Manicaland Province (Schaefer *et al.*, 2017). Another study by Oseni *et al.* (2018) investigated the sub-regional changes in HIV prevalence in Zimbabwe (Oseni *et al.*, 2018), while Gwitira *et al.* (2018) tested for the co-existence of the geographical distribution of HIV and malaria in Zimbabwe (Gwitira *et al.*, 2018). Cuadros *et al.*, (2019) constructed high-resolution maps of intra-national estimates of HIV prevalence in Zimbabwe based on spatial variables (Cuadros *et al*., 2019) and Manyangadze *et al.* (2021) looked at socio-economic factors spatially associated with HIV incidence in Zimbabwe (Manyangadze *et al.*, 2021).

In Zimbabwe, the HIV epidemic has been found to be heterogeneous, varying by province, with Matabeleland South having the highest burden (Gonese *et al.*, 2020). Given a set of covariates that drive the HIV epidemic in Zimbabwe, this means that because of the heterogeneous nature of the epidemic, the responses to these covariates across the country would be different. Although significant work has been done in Zimbabwe regarding spatial analysis, besides spatially modelling the HIV epidemic in terms of prevalence and at the province level, not much has been done in modelling the epidemic in terms of incidence and at the district level. Understanding the HIV epidemic and its spatial variation is essential as it allows the implementation of policies targeted explicitly at high risks areas and groups. The use of geospatial analytical methods is useful in achieving this goal. However, among the limited but growing uses of geospatial analysis in Africa, moreover in Zimbabwe, very few studies have analysed the spatiotemporal variation in the HIV epidemic. This knowledge gap hinders our understanding of the specific factors driving the local transmission dynamics and impedes the design of targeted interventions.

To address this gap, our study aimed to examine the relationship between district patterns of HIV incidence and a range of covariates in Zimbabwe over a 10-year period (2005-2015) using the demographic health survey data and utilising spatial analysis methods. By considering both individual and contextual factors, we seek to identify the key determinants associated with the heterogeneity in HIV transmission at the district level. Moreover, we aimed to explore whether the relationships between these covariates and HIV incidence vary across different districts in Zimbabwe.

Materials and Methods

Data source

The source of data was the Zimbabwe Demographic and Health Survey (ZDHS) conducted in 2005-06 (Zimbabwe Central Statistical Office and Macro International, 2007), 2010-11(Zimbabwe National Statistics Agency - ZIMSTAT and ICF International, 2012) and 2015 (Zimbabwe National Statistics Agency and ICF International, 2016). Enrolment of individuals in the ZDHS involved a two-stage sampling procedure in selecting households. The first stage involved selecting enumeration areas (EAs) and the second stage involved randomly selecting households from the EAs. Full details of the survey methodology are described elsewhere (Birri Makota & Musenge, 2022). A representative probability sample consisting of 10,800, 10,828 and 11,196 households was selected for the 2005-06, 2010-11 and 2015 ZDHS databases, respectively. In the 2005-06 ZDHS database, there were 16,082 records for individuals aged between 15 and 49 years, comprising of 7,175 (44.6%) males and 8,097 (55.4%) females. Similarly, the 2010-11 ZDHS database contained 16,651 records for individuals aged 15-49 years, including 7,480 (44.9%) males and 9,171 (55.1%) females. Finally, in the 2015 ZDHS database, there were 18,351 records for individuals aged 15-49 years, consisting of 8,396 (45.8%) males and 9,955 (54.2%) females. The individual records for males and females of each survey year were appended and these datasets were then merged to the HIV datasets. Lastly, the datasets of Global Positioning System (GPS) coordinates of all the EAs were merged to the HIV individual records, which means that all individuals were assigned the location of the centroid of their respective survey EA. After assigning each individual to the respective EA with GPS coordinates, each individual was placed in their respective districts. Zimbabwe is comprised of 60 districts (Figure 1).

Study measures

The outcome variable for our analysis was HIV incidence. This incidence for each district was estimated using the catalytic model. Details of the estimation procedure are explained elsewhere (Birri Makota & Musenge, 2022). In the data exploratory analysis, a stepwise logistic regression approach was performed using the svy:swaic command (StataCorp., 2017) in STATA SE. version 15.1 statistical software. This method was employed to identify variables that potentially influenced the risk of HIV. Only factors that showed significant associations with HIV, as determined by stepwise survey logistic regression (Birri Makota & Musenge, 2019), were considered for further analysis. The independent variables

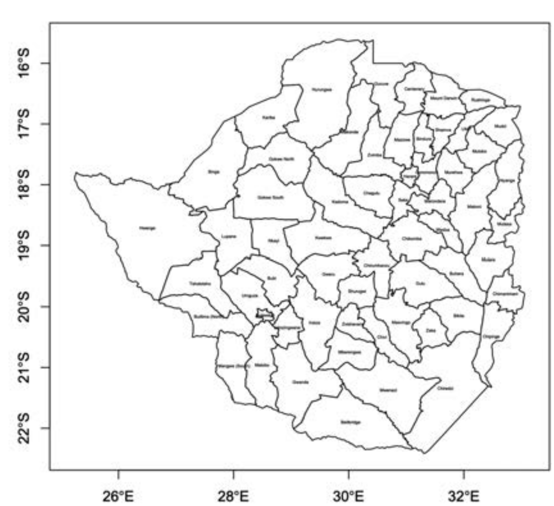

Figure 1. Map of Zimbabwe showing the 60 districts.

included age, sex, marital status, condom use, age of partner, STI in the past year, first sexual debut before the age of 15 years, wealth index, age of household head, multiple sex partners, times away from home, highest level of education, type of earnings and employment status.

Exploratory data analysis

Hotspot cluster and outlier analysis

Hotspot Analysis was performed using the Getis-Ord Gi* statistic (Mitchell, 1999), which identifies statistically significant hotspots and coldspots using a set of weighted features (Manepalli *et al.*, 2011). Each district was assigned a Z-score, which identified the intensity of the hotspot and coldspot zones for the HIV incidence. The index of the Getis-Ord Gi* statistic was calculated as follows:

$$G_i^* = \frac{\sum_j^n w_{ij} y_j - \bar{y} \sum_j^n w_{ij}}{s\sqrt{\frac{n\sum_j^n w_{ij}^2 - \left(\sum_j^n w_j\right)^2}{n-1}}} \quad \text{(Eq. 1)}$$

where y_j is the HIV incidence for district *j*; w_{ij} the spatial weights between *i* and *j*: *n* the total number of districts; $\bar{y}$ the mean; and *s* the standard deviation (SD).

To produce smooth surfaces of HIV incidence for visualisation and data generation at the district level, HIV incidence interpolation maps were created through the Empirical Bayesian Kriging (EBK) (Anselin *et al.*, 2006). Through repeated EBK simulations, accounts for the error in semi variogram estimation were created (Gribov & Krivoruchko, 2020). Advantages of the EBK include predicting more accurate standard errors than other Kriging methods and allowing the accurate prediction of moderately non-stationary data (Gribov & Krivoruchko, 2020). Contrary to the hotspot analysis, which identifies groups within an area, the cluster and outlier analysis identifies groupings according to criterion of proximity. Cluster and outlier analysis were performed using local Moran's *I* (Anselin, 1995). These techniques were all performed using software issued by the Environmental Systems Research Institute (ESRI) using ArcGIS Pro, version 3 (ESRI, 2023).

Spatial statistical models

Exploratory regression

To identify independent variables associated with HIV infection, we first conducted an exploratory regression (ER) of the 14 variables mentioned above for each survey year separately. The ER was performed in ArcGIS Pro version 3 using the ER tool, a data mining tool that finds all possible combinations of the explanatory variables that will pass the necessary ordinary least regression (OLS) diagnostics (Hashim *et al.*, 2019). The advantage of ER in ArcGIS above ordinary stepwise regression is the ability to find models that meet all the requirements and assumptions of the OLS rather than just searching for models with the highest adjusted R^2. For the 2005-2006 ZDHS, the significant risk factors obtained from ER included an STI in the past year for both married and single persons; for the 2010-11 ZDHS, they additional factors were being financially better off according to the ZDHS wealth index, both for singles and females with partners five or more years older. The significant factors from ER of the 2015 ZDHS included females with partners five or more years older than them, having an STI in the past year, having multiple sex partners and also being financially better off according to the ZDHS wealth index. These significant variables were then run in the OLS model, with residuals checked for spatial autocorrelation.

OLS regression – Global regression

To determine the global relationship between HIV incidence and the significant covariates, the OLS model was fitted for each survey. The three OLS models fitted for the 2005-06 survey year (Eq. 2), the 2010-11 survey year (Eq. 3) and the 2015 survey year (Eq. 4) were as follows:

$$HIV = \beta_0 + \beta_1 PropSTI_i + \beta_2 PropMarried_i + \beta_3 PropRicher_i + \varepsilon_i \quad \text{(Eq. 2)}$$

$$HIV = \beta_0 + \beta_1 PropRicher_i + \beta_2 PropSingle_i + \beta_3 OlderPartner_i + \varepsilon_i \quad \text{(Eq. 3)}$$

$$HIV = \beta_0 + \beta_1 PropOldpartner_i + \beta_2 PropSTI_i + \beta_3 PropRicher_i + \beta_4 PropSexPartners_i + \varepsilon_i \quad \text{(Eq. 4)}$$

where HIV is the estimated incidence; β_0 the intercept; β_i regression coefficients; *i* the 60 districts; *PropSTI*: the weighted proportions of individuals who had an STI in the past 12 months; *PropMarried*: the weighted proportions of individuals married; *PropRicher*: the weighted proportions of individuals financially better off; *PropSingle*: the weighted proportions of individuals single; *PropOldpartner*: the weighted proportions of individuals with partners five or more years older; and *PropSexpartners*: weighted proportions of individuals who have multiple sex partners.

Considering that OLS assumes stationarity (*i.e.*, the association between response and predictor variable does not vary spatially). Therefore, to test for stationarity in the spatial pattern, Moran's *I* statistics for spatial autocorrelation (Aswi *et al.*, 2021) was used as follows:

$$I = \frac{n\sum_i^n \sum_j^n w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\left(\sum_i^n \sum_j^n w_{ij}\right)\sum_i (y_i - \bar{y})^2} \quad \text{(Eq. 5)}$$

Moran's *I* is a measure of spatial autocorrelation in an area, which explains the degree of dependence between values of a variable at different locations. In addition, we tested for multicollinearity using the variance inflation factor (VIF) and local variation using the Koenker (BP) statistic (Ahmad *et al.*, 2021). OLS was performed in ArcGIS Pro version 3.1. Using the same set of variables used in the OLS, we performed the geographically weighted regression (GWR) and multi-scale geographically weighted regression (MGWR) to account for nonstationarity between response (HIV incidence) and predictor variables (Fotheringham *et al.*, 2017). The GWR and MGWR were performed using the MGWR software which is also integrated in the *mgwr* Python package (version 2.2) (Oshan *et al.*, 2019).

The GWR model

Considering that the covariates' predictability capacity differs from one cluster to the next, these cluster variations can be identified through the use of GWR (Wheeler & Páez, 2010). The GWR assesses whether the association between HIV incidence and covariates vary across space. The GWR creates an equation for each district while using data from nearby features; hence the coefficient takes different values for each district. The maps created through GWR of the coefficients associated with each explanatory variable provide guidelines for targeted interventions within different districts. The GWR model can be expressed as follows:

$$y_i = \beta_0(u_i v_i) + \beta_1(u_i v_i)x_{i1} + \beta_2(u_i v_i)x_{i2} + \ldots + \beta_j(u_i v_i)x_{ij} + \varepsilon_i = \sum_{j=0}^{n} \beta_j(u_i v_i)\, x_{ij} + \varepsilon_i \quad \text{(Eq. 6)}$$

where y_i are observations of HIV incidence; $u_i v_i$ are the longitude and latitude geographical coordinates for the *ith* observation; β_j $(u_i v_i)$ athe parameters for x_{ij} describing the relationship around geographical locations $(u_i v_i)$ and ε_i is the error term.

The MGWR model

The GWR constrains the local (district) relationships within each model to vary at the same scale. However, the MGWR allows the response variable (HIV incidence) and the different covariates to vary at different spatial scales (Fotheringham *et al.*, 2017). This means that the bandwidths indicating the data-borrowing range can vary across parameter surfaces. Therefore, the equation of the MGWR differs slightly from the GWR by introducing the bandwidth:

$$y_i = \beta_{bw0}(u_i v_i) + \beta_{bw1}(u_i v_i)x_{i1} + \beta_{bw2}(u_i v_i)x_{i2} + \ldots + \beta_{bwj}(u_i v_i)x_{ij} + \varepsilon_i = \sum_{j=0}^{n} \beta_{bwj}(u_i v_i)\, x_{ij} + \varepsilon_i \quad \text{(Eq. 7)}$$

where the bandwidth used for calibration of the jth conditional relationship is represented by *bwi* in β_{bwj}.

Results

Figure 2 shows the spatial distribution of HIV incidence over the 10 years under study. The highest HIV district incidence rates recorded were 0.029, 0.025 and 0.026 for the years 2005-2006, 2010-2011 and 2015, respectively. High HIV incidence rates were observed in 11 (18.3%), 8 (13.3%) and 9 (15.0%) districts in the years 2005-2006, 2010-2011 and 2015, respectively. The level of infection in most of the districts with high HIV incidence rates in 2005-2006 and 2010-2011 decreased in 2015.

Hotspots and clusters

The Getis-Ord Gi* statistic was used to reveal spatial patterns and statistically significant clustering of districts of the HIV incidence into hotspots and coldspots in different survey years under consideration. In 2005-2006, hotspots for HIV incidence with 99% confidence were observed in 11 (18.3%) districts, most of them located in the southern region of Zimbabwe (Figure 3a). No significant coldspots at the level of confidence were observed in 2005-2006; only one at 90% confidence. In 2010-2011, hotspots with 99% confidence were observed in districts in the western parts of Zimbabwe (Figure 3b). In 2015, there were three such hotspots remaining, which were all in Matabeleland North Province; Tsholotsho, a district in this province, remained a hotspot at the 99% level throughout the decade under observation. Based on the HIV hotspot analysis depicted in Figure 3a-c, it can be noticed that these hotspots decreased significantly in Zimbabwe over time.

Figure 3d-f shows that the HIV incidence clusters were delimited within four zones. The High-High clusters, areas characterised by high HIV incidence risk surrounded by districts with similar risk, were mostly located in districts of Matabeleland South Province in 2005-2006. The presence of High-High clusters in districts of this province continued into the 2010-2011 survey as well as in the 2015 one.

Interpolation of HIV incidence hotspots is shown in Figure 3(g-i) diagrams. Areas with high HIV incidence are shown in red (hotspots), while areas with low HIV incidence are shown in blue (coldspots). From Figure 3(g-i), continuous surfaces of HIV incidence also confirmed and spatially unpacked (disaggregating the data to understand the specific spatial patterns) district and regional HIV incidence variations in Zimbabwe. Figure 3(g-i) captured the general intensification of HIV incidence and spatial extent from 2005 to 2015. It can be noticed that the southern and western parts of the country are characterized by high HIV incidence rates, whereas the eastern and northern regions display coldspots.

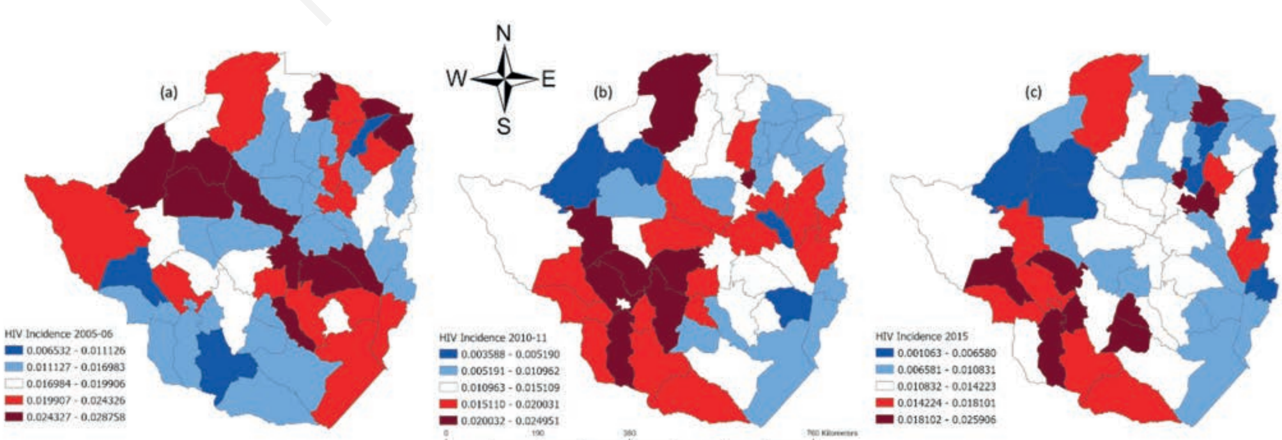

Figure 2. Spatial Variation of HIV incidence per survey year. Red colour shades indicate districts with the high HIV incidence and blue colour shades the districts with low HIV incidence. Strength of presence is indicated by depth of colour.

OLS results

OLS was used to determine the best combination of variables to optimise the prediction of HIV incidence for each survey year. We therefore produced three models for each survey year and the final variables used in all three models were selected through a multi-stage statistical process from an initial pool of 14 variables drawn from literature and based on availability at the survey year and district level. The first step included ER that was done on all the 14 variables on the three survey years separately. The ER yielded three variables for the 2005-06 data (STI, married and single), three for

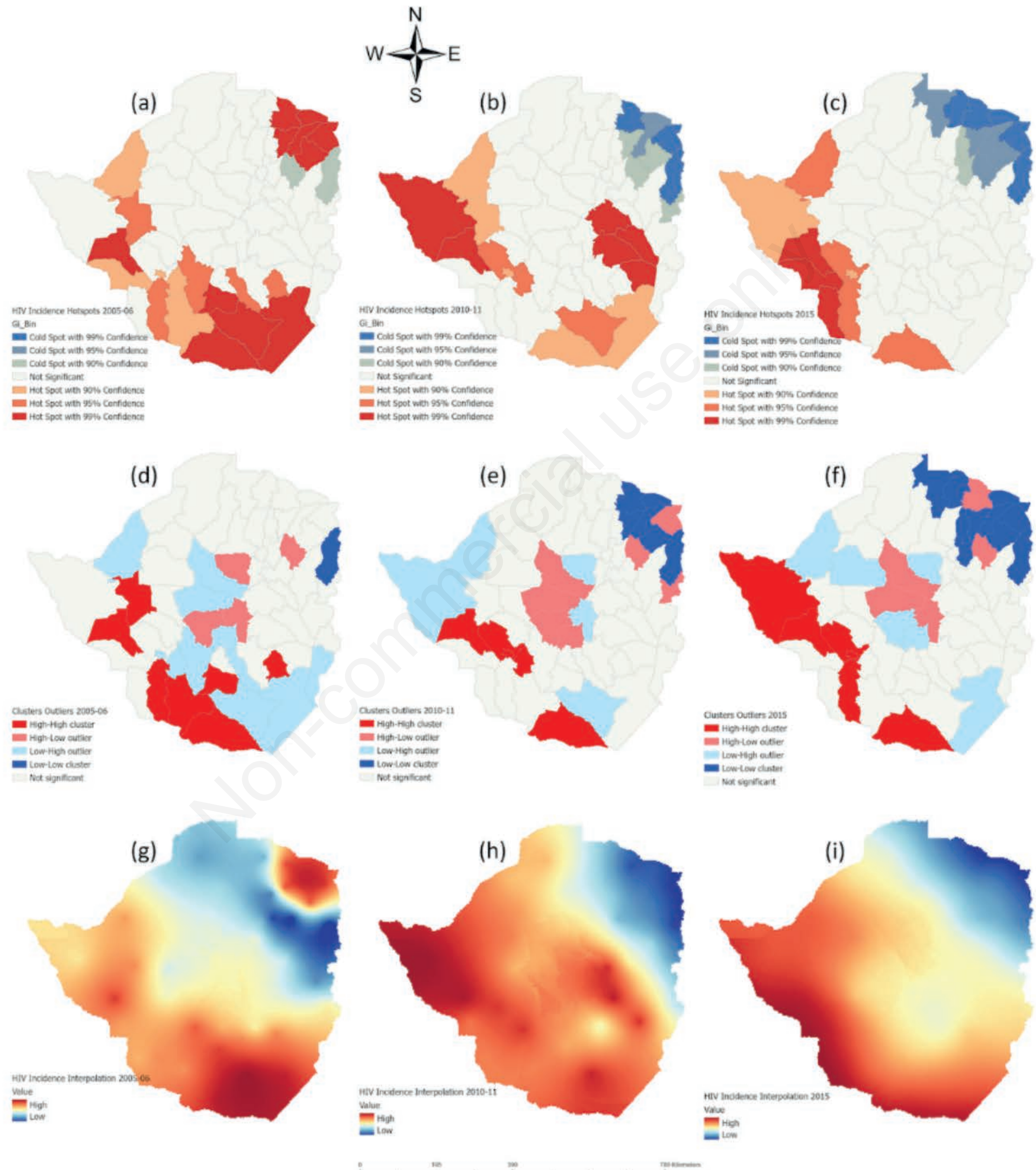

Figure 3. HIV incidence represented by hotspots and clusters. (a-c) hotspots, (d-f) clusters and (g-i) interpolation per survey year. Hotspots in the maps are shown in red shades and coldspots in blue shades; non-shaded areas are statistically not significant; High-High clusters are shown in red shades and Low-Low ones in blue shades. Strength of presence is indicated by depth of colour.

Table 1. Ordinary least squares regression (OLS) results.

Variable	Coefficient	Standard error	p	VIF	Moran's *I* p	Koenker (BP) p	Jarque-Bera p
2005-06							
Intercept	-0.025	0.003	<0.001				
STI	0.032	0.015	0.03	1.02	-0.01	14.33	0.91
Not using condom	0.020	0.005	0.02	1.43	(0.59)	(0.002)	(0.63)
Married	0.039	0.005	<0.001	1.46			
2010-11							
Intercept	0.001	0.001	0.37				
Single	0.009	0.004	0.01	1.45	-0.03	12.31	0.29
Financially better off	0.019	0.004	<0.001	2.60	(0.40)	(0.006)	(0.86)
Older partner >5 years	0.011	0.003	0.01	2.17			
2015							
Intercept	-0.002	0.001	0.02				
Multiple sex partners	0.013	0.004	<0.001	4.17	-0.03	10.20	4.07
STI	0.043	0.019	<0.001	1.02	(0.26)	(0.037)	(0.13)
Financially better off	0.008	0.003	0.03	1.06			
Older partner >5 years	0.010	0.004	<0.001	4.17			

VIF, variance inflation factor; STI, sexually transmitted infections.

the 2010-11 data (single, financially better off, and older partner >5 years) and four for the 2015 data (multiple sex partners, STI, financially better off and older partner >5 years). These variables were then used for the OLS producing the results displayed in Table 1. As shown in this Table, all variables for the three survey years were statistically significant. Based on the OLS results, we checked for multicollinearity, residual stationarity, and global spatial autocorrelation. Results from the spatial autocorrelation indicated that, all the three models' residuals were randomly distributed with p-values of 0.59, 0.40 and 0.26 of Moran's *I* for the 2005-2006, 2010-2011 and 2015 datasets, respectively. Since these p-values were far from statistically significant, the structure of the residuals was deemed to be random, and the model properly specified. It was also observed that all the variables for all the survey years under study had VIF values <5, which indicated that there was low multicollinearity in the models. The Jarque-Bera statistic (Huang *et al.*, 2022) was not statistically significant for any of the models meaning that the models were not biased, and key variables were selected. Lastly, the Koenker (BP) statistic was significant for all models in all survey years with p-values of 0.002, 0.006 and 0.037 for the 2005-2006, 2010-2011 and 2015 respectively. This meant that the residuals of the errors from the OLS were non-stationary and the GWR or MGWR were the best models to model the spatial variability. The spatial variation of the selected variables is presented as choropleth maps in the Appendix file.

Geographical/spatial variation of HIV incidence

As observed in the results presented in Table 2, the MGWR model is a better fit based on the adjuster R^2 and Akaike's corrected information criterion (AICc) values. In addition, the MGWR model explained 81.5%, 86.5% and 80.2% of the total variation in HIV incidence in the 2005-2006, 2010-2011 and 2015 models, respectively, which was an improvement from the OLS and GWR approaches. For additional comparison between MGWR and GWR, see Appendix file with spatial maps for both models.

Figure 4 shows how the coefficients for STI, non-use of condom and married status for the 2005-2006 survey year are spatially associated with HIV incidence. These variables were all positive predictors of high HIV incidence in 2005-2006, indicating that HIV incidence rates increases at the same rate as the percentage of individuals with an STI, who do not use condoms and are married increased. As observed in Figure 4, the impact of these variables has been strong in the eastern parts of Zimbabwe, *i.e.*, Mashonaland Central, Mashonaland East and Manicaland provinces.

Figure 5 shows how the coefficients for those financially better off, being single and being female with an older partner >5 years in 2010-2011 survey year are spatially associated with HIV incidence. As shown by districts in the figure, being financially better off in districts situated in Matabeleland North was a strong predictor of HIV incidence. As the percentage of single individuals increased in Masvingo Province, the HIV incidence also increased by (24-25%) as shown in Figure 5b indicating that this variable also was a strong predictor of HIV incidence in this province.

Table 2. Comparison of regression results.

Survey year	Model	AICc	Adjusted R^2 (%)
2005-06	OLS	145.6	77.0
	GWR	90.0	77.0
	MGWR	82.7	81.5
2010-11	OLS	135.2	72.5
	GWR	103.6	73.2
	MGWR	97.1	86.5
2015	OLS	140.3	70.9
	GWR	102.1	70.9
	MGWR	99.3	80.2

AICc, Akaike's corrected information criterion; Adjusted R^2, adjusted coefficient of determination; OLS, ordinary least regression, GWR, geographically weighted regression; MGWR, multi-scale geographically weighted regression.

Females having older partners were strongly associated with high HIV incidence and the areas mostly affected were Binga, Kariba, Gokwe North and Gokwe South districts as shown in Figure 5c.

Figure 6 shows that the coefficients for having multiple sex partners, STI, being financially better off and females with older partners >5 years in 2015 survey year are spatially associated with higher HIV incidence. All districts in Matabeleland North showed that a high percentage increase for individuals having multiple sex partners and STI had a very strong positive association with HIV incidence. Being financially better off had a strong impact on HIV incidence in the districts of Masvingo province, mainly Chiredzi, Mwenezi, Mberengwa, Zvishavane and Chivi. Females having an older partner were significantly associated with a high HIV incidence in districts of Lupane and Matabeleland North. Given that STI and multiple sex partners have a strong positive association with HIV incidence in the same districts as shown in Figure 6a-b, interventions encouraging partner notification when one has an STI especially with multiple sex partners would therefore help reduce asymptomatic STIs in the districts, which, in turn, should reduce HIV incidence rates.

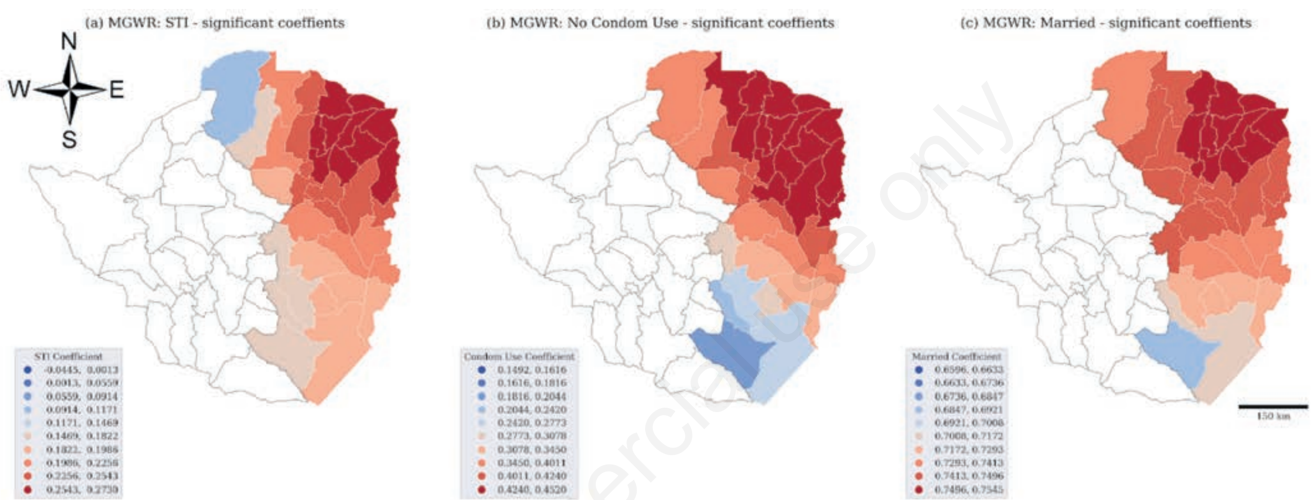

Figure 4. HIV coefficient estimates for three variables (sexually transmitted infection in the past 12 months; non-use of condom; and married status) based on multi-scale geographically weighted regression for the 2005-2006 survey year. Areas with ±1.96 significance represented in red shades show where the coefficients are large and locations where the explanatory variables are strong predictors of high HIV incidence. Red colour shades indicate districts with the high HIV incidence and blue colour shades the districts with low HIV incidence. Strength of presence is indicated by depth of colour.

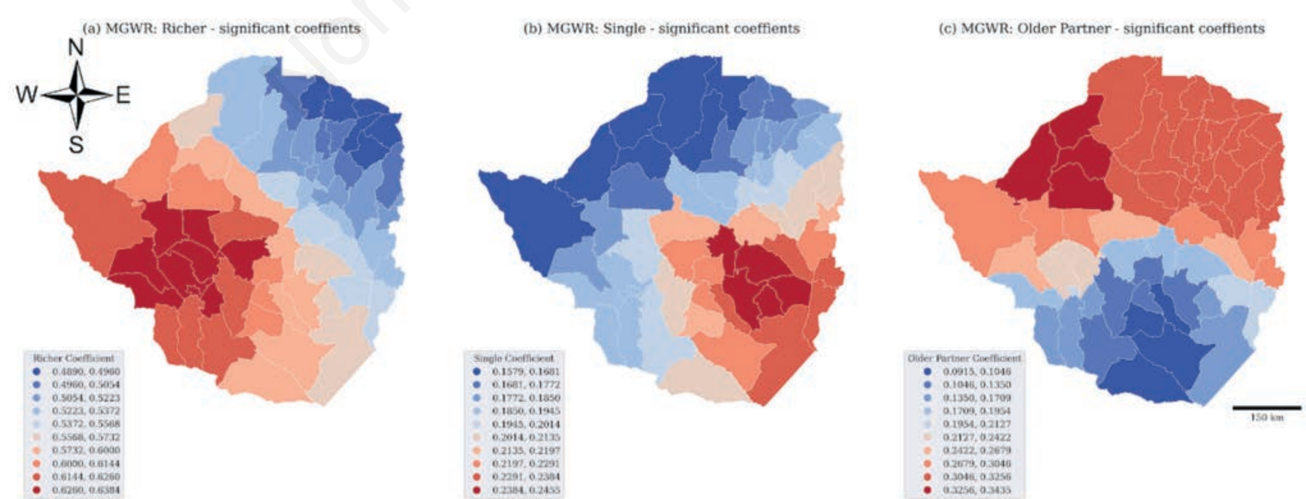

Figure 5. HIV coefficient estimates of for three variables (financially better off; being single; been together with an older partner >5 years). Areas with ±1.96 significance represented in red shades show where the coefficients are large and locations where the explanatory variables are strong predictors of high HIV incidence. Red colour shades indicate districts with the high HIV incidence and blue colour shades the districts with low HIV incidence. Strength of presence is indicated by depth of colour.

Discussion

This study explored the spatial heterogeneity and spatial predictors of HIV incidence over a decade using spatial analytic methods. The hotspot areas (local clusters) were observed in the southern and western parts of Zimbabwe during this time. Notably, in 2005-2006, the southern parts of Zimbabwe had many hotspot districts. These findings can be corroborated by reports stating that HIV prevalence and incidence are high in the southern parts of Zimbabwe (Gonese *et al.*, 2020; Makurumidze *et al.*, 2020; Moyo *et al.*, 2017). Zimbabwe's southern region shares borders with Botswana and South Africa. Botswana and South Africa have been reported in literature to have the highest burden of HIV in Sub-Saharan Africa (Dwyer-Lindgren *et al.*, 2019). Therefore, we can hypothesise that the districts in the southern parts of Zimbabwe have the highest burden of HIV since they border with nations (South Africa and Botswana) that have also recorded high HIV incidences. One of the provinces situated in the southern region is Matabeleland North, which constitutes of Tsholotsho, Lupane, Nkayi, Bubi and Umguza districts. From our findings from the MGWR, we observed that Matabeleland North's HIV incidence rates are driven by the wealth index (Figure 5a), multiple sex partners (Figure 6a), STI (Figure 6b) and females with older partners (Figure 6d). An example of one of the districts with the highest burdens of HIV in the southern region of Zimbabwe is Tsholotsho, a district situated in Matabeleland North reported to have the very highest HIV burden (Moyo *et al.*, 2017). It has been reported that the key drivers of HIV in Tsholotsho include high gender-based violence (GBV), low condom use, low-risk perception and unequal power relations between males and females (Lumbidzani, 2022). In addition, due to its proximity to the border, this district experiences high levels of migration, which often leads to a significant number of couples separating. Further research into why Tsholotsho has remained a hotspot area for over a decade is required in order to tailor interventions that effectively transform this district's HIV burden narrative.

This study has highlighted that the proportion of wealthier individuals is positively associated with high HIV incidence rates for all three survey years. The proportion of financially better off individuals was obtained from the variable wealth index from the demographic health survey (DHS), which is a composite measure of a household's cumulative living standard (Zimbabwe Central Statistical Office and Macro International, 2007, Zimbabwe National Statistics Agency and ICF International, 2016, Zimbabwe National Statistics Agency - ZIMSTAT and ICF International, 2012). A study by Fox (2012) indicated that in wealthier countries, poor individuals had a higher risk of HIV infection, whereas, in poorer countries, individuals who had more wealth were more likely to be infected by HIV (Fox, 2012). Zimbabwe cannot be described as a wealthy nation; therefore, the conclusion that financially better off individuals are at a higher risk of HIV infection is applicable. Another study revealed that, in most cases, the HIV burden among adults increased along with wealth (Mishra *et al.*, 2007). These two findings support our finding that areas with a higher proportion of better off individuals have a higher HIV incidence rate. However, another study by Gaumer *et al.* (2021) suggested that the DHS survey samples may contain an inflated number of HIV-positive individuals surviving in the wealthier group, thus living longer than poorer individuals, giving the impression that the HIV burden is higher in the group of people who are better off (Gaumer *et al.*, 2021). Whether or not the DHS sampling introduces a bias, policies targeting areas with high wealth inequalities are required.

This study observed that proportions of females with partners >5 years older than them were positively associated with an increase in HIV incidence in 2010-2011 and 2015. Unfortunately, we could not test the same covariate in 2005-2006 as this variable was not yet part of the questions asked in the DHS questionnaire. Age-disparate or age-mixing patterns refer to sexual relationships where the age difference between partners is five years or greater (Mabaso *et al.*, 2021). A nation population-based household survey conducted in South Africa noted that females who resided in rural areas were more likely to have a partner who was ≥5 years or greater than them compared to females who resided in urban areas (Mabaso *et al.*, 2021). Our findings support this notion as districts with a high proportion of females with partners >5 years older are rural districts. Additionally, in rural areas, especially in poor rural areas, lack of employment, education and general economic depri-

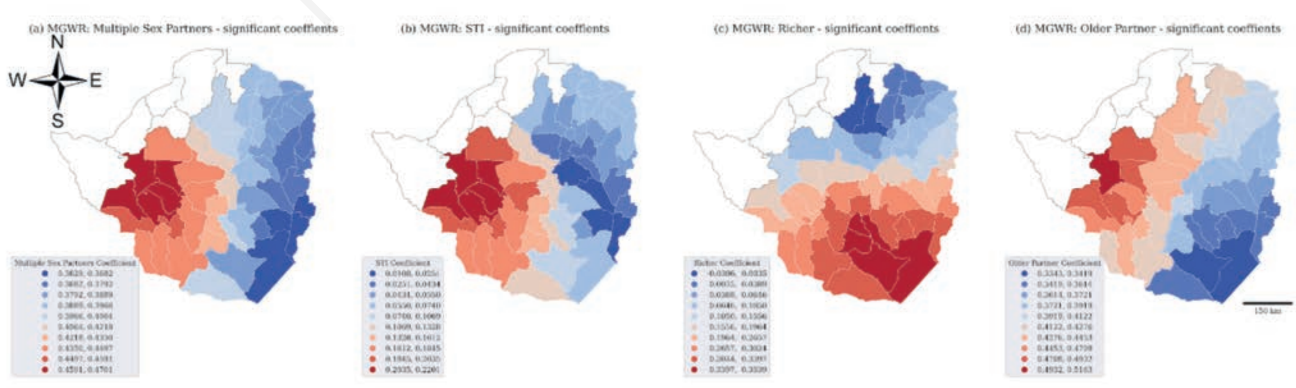

Figure 6. HIV coefficient estimates for four variables (having multiple sex partners; had a sexually transmitted infections in the past 12 months; being financially better off; and having an older partner >5 years based on multi-scale geographically weighted regression for the 2015 survey year with. Areas with ±1.96 significance represented in red shades show where the coefficients are large and locations where the explanatory variables are strong predictors of high HIV incidence. Red colour shades indicate districts with the high HIV incidence and blue colour shades the districts with low HIV incidence. Strength of presence is indicated by depth of colour.

vation leads to women seeking older partners for financial security (Leclerc-Madlala, 2008; Wabiri *et al.*, 2016). More investigations are required to understand the HIV transmission dynamics between living in rural areas and having partners >5 years older. This investigation will warrant a tailored response to those specific areas thereby strengthening already existing HIV programmes.

Condoms are one of the most effective ways of reducing STIs and HIV transmission rates, but only if they are used properly and consistently. Therefore, promotion and free distribution of condoms in the Eastern parts of Zimbabwe focusing on married individuals would decrease HIV transmission (Figure 4a-c). Additionally, areas with high burdens of STIs, sexual reproductive health services should be free and easily accessible. As given that the Eastern parts of Zimbabwe are consistently affected by three variables, *i.e.*, sexually transmitted infection in the past 12 months; non-use of condom; and married status (Figure 4), integration of HIV/STI services into contraceptive services would greatly reduce HIV incidence. In addition, targeted structural interventions aimed at tackling social and economic constraints across all wealth quantiles should improve HIV burdens.

Study limitations

Although the study contributes significantly to advancing spatial analysis of the HIV epidemic in Zimbabwe, it has limitations. First, since this was a secondary data analysis from a cross-sectional study, our results are affected by recall bias and social desirability since the covariates were self-reported. Secondly, since this was a cross-sectional study design, we cannot infer causality. Since we estimated HIV incidence for each district from HIV prevalence, we noted that the data were not geared for district-level analysis as some districts had very low HIV prevalence rates, which could have biased the final estimated value. However, since the results were corroborated by other studies, we believe that the HIV incidence estimates were not far off from the true estimates.

Overall, while the absence of district-level data poses challenges, researchers can employ innovative approaches, data imputation techniques, surrogate indicators, and qualitative research to gain insights into district-level HIV transmission dynamics. Collaborative efforts to improve data collection, data sharing, and standardisation of district-level data should further strengthen future studies and enable more accurate and comprehensive spatial analysis at the local level.

Conclusions

Regardless of the limitations, the study found relationships between covariates and HIV incidence over space and disseminated the message that different factors vary from district to district over time. Our findings also indicate the need to zoom-in on districts like Tsholotsho, which consistently have a high burden of HIV over time. Interventions that promote partner notification, e.g., with respect to people with STI, especially with multiple sex partners, should contribute to a reduction of asymptomatic STIs in the districts, something expected to reduce HIV incidence. We encourage the implementation of interventions that address and improve access to education for women, and promote economic empowerment, allowing women to have greater control over their sexual decisions and health thereby reducing their vulnerability to HIV infection. This study recommends engaging and involving local communities in planning, implementing, and monitoring HIV prevention and treatment programmes. This can be achieved by introducing community-based approaches, such as peer education and support groups, which can effectively reach individuals in remote or marginalised areas where uptake of healthcare services may be limited. We also recommend strengthening the collection and analysis of district-level HIV data to inform evidence-based interventions. Regular monitoring and evaluation of HIV incidence, prevalence, and risk factors at the district level can help identify hotspots and design targeted interventions accordingly. The findings presented here should be useful to policymakers for the purposes of resource allocation in the context of public health programmes.

References

Ahmad I, Dar MA, Fenta A, Halefom A, Nega H, Andualem TG, Teshome A. 2021. Spatial Configuration Of Groundwater Potential Zones Using Ols Regression Method. J Afr Earth Sci 177;104147.

Anselin L, 1995. Local Indicators Of Spatial Association—Lisa. Geograph Anal 27;93-115.

Anselin L, Lozano N, Koschinsky J. 2006. Rate Transformations And Smoothing. Urbana, 51, 61801.

Aswi A, Cramb S, Duncan E, Mengersen K. 2021. Detecting Spatial Autocorrelation For A Small Number Of Areas: A Practical Example. J Physics: Conference Series, 2021;012098.

Aturinde A, Farnaghi M, Pilesjö P, Mansourian A, 2019. Spatial Analysis Of Hiv-Tb Co-Clustering In Uganda. Bmc Infect Dis 19;1-10.

Awaidy SA, Ghazy RM, Mahomed O. 2023. Progress Of The Gulf Cooperation Council (Gcc) Countries Towards Achieving The 95-95-95 Unaids Targets: A Review. J Epidemiol Global Health, 1-10.

Birri Makota R, Musenge E, 2019. Factors Associated With HIV Infection In Zimbabwe Over A Decade From 2005 To 2015: An Interval-Censoring Survival Analysis Approach. Front Public Health 7;262.

Birri Makota R, Musenge E, 2022. Estimating Age Specific Prevalence And Force Of Infection In Zimbabwe Using Combined Cross-Sectional Surveys From 2005-2015. Front Epidemiol 2;59.

Bulstra CA, Hontelez JA, Giardina F, Steen R, Nagelkerke NJ, Bärnighausen T, De Vlas SJ, 2020. Mapping And Characterising Areas With High Levels Of HIV Transmission In Sub-Saharan Africa: A Geospatial Analysis Of National Survey Data. Plos Medicine 17:E1003042.

Coburn BJ, Okano JT, Blower S, 2013. Current Drivers And Geographic Patterns Of Hiv In Lesotho: Implications For Treatment And Prevention In Sub-Saharan Africa. BMC Medicine 11;1-11.

Cuadros DF, Li J, Branscum AJ, Akullian A, Jia P, Mziray EN, Tanser F, 2017. Mapping The Spatial Variability Of Hiv Infection In Sub-Saharan Africa: Effective Information For Localized HIV Prevention And Control. Sci Rep 7;9093.

Cuadros DF, Li J, Mukandavire Z, Musuka GN, Branscum AJ, Sartorius B, Mugurungi O, Tanser F, 2019. Towards Unaids Fast-Track Goals: Targeting Priority Geographic Areas For HIV Prevention And Care In Zimbabwe. AIDS 33;305-314.

Dwyer-Lindgren L, Cork MA, Sligar A, Steuben KM, Wilson KF, Provost NR, Mayala BK, Vanderheide JD, Collison ML, Hall JB, Biehl MH, Carter A, Frank T, Douwes-Schultz D, Burstein R, Casey DC, Deshpande A, Earl L, El Bcheraoui C, Farag TH, Henry NJ, Kinyoki D, Marczak LB, Nixon MR, Osgood-Zimmerman A, Pigott D, Reiner RC, Ross JM, Schaeffer LE, Smith DL, Davis Weaver N, Wiens KE, Eaton JW, Justman JE, Opio A, Sartorius B, Tanser F, Wabiri N, Piot P, Murray CJL, Hay SI, 2019. Mapping HIV Prevalence In Sub-Saharan Africa Between 2000 And 2017. Nature 570;189-193.

ESRI 2023. Arcgis Pro. 3.1 Ed. Redlands, Ca: Environmental Systems Research Institute.

Fotheringham AS, Yang W, Kang W, 2017. Multiscale Geographically Weighted Regression (Mgwr). Ann Am Ass Geograph 107;1247-1265.

Fox AM, 2012. The HIV–Poverty Thesis Re-Examined: Poverty, Wealth Or Inequality As A Social Determinant Of HIV Infection In Sub-Saharan Africa? J Biosoc Sci 44;459-480.

Gaumer G, Sherafat-Kazemzadeh R, Jordan M, Nandakumar A, 2021. Wealth And Wealth Inequality In Adult HIV Prevalence. J Global Health Rep 4;E2020105.

Gonese E, Musuka G, Ruangtragool L, Hakim A, Parekh B, Dobbs T, Duong YT, Patel H, Mhangara M, Mugurungi O. 2020. Comparison Of HIV Incidence In The Zimbabwe Population-Based HIV Impact Assessment Survey (2015–2016) With Modeled Estimates: Progress Toward Epidemic Control. AIDS Res Human Retrovir 36;656-662.

Gribov A, Krivoruchko K. 2020. Empirical Bayesian Kriging Implementation And Usage. Sci Total Environ 722;137290.

Gwitira I, Murwira A, Mberikunashe J, Masocha M, 2018. Spatial Overlaps In The Distribution Of Hiv/Aids And Malaria In Zimbabwe. BMC Infect Dis 18;1-10.

Hashim H, Wan Mohd W, Sadek E, Dimyati K. 2019. Modeling Urban Crime Patterns Using Spatial Space Time And Regression Analysis. Int Arch Photogrammetry Remote Sensing Spatial Inform Sci 42;247-254.

Huang J, Mao X, Deng H, Liu Z, Chen J, Xiao K, 2022. An Improved Gwr Approach For Exploring The Anisotropic Influence Of Ore-Controlling Factors On Mineralization In 3d Space. Nat Resour Res 1-16.

Jahagirdar D, Walters MK, Novotney A, Brewer ED, Frank TD, Carter A, Biehl MH, Abbastabar H, Abhilash E, Abu-Gharbieh E, 2021. Global, Regional, And National Sex-Specific Burden And Control Of The Hiv Epidemic, 1990–2019, For 204 Countries And Territories: The Global Burden Of Diseases Study 2019. Lancet HIV 8;E633-E651.

Kharsany AB, Karim QA, 2016. Hiv Infection And Aids In Sub-Saharan Africa: Current Status, Challenges And Opportunities. Open Aids J 10;34.

Leclerc-Madlala S. 2008. Age-Disparate And Intergenerational Sex In Southern Africa: The Dynamics Of Hypervulnerability. AIDS 22;S17-S25.

Lumbidzani D. 2022. Why Tsholotsho Has The Highest HIV Burden. Chronicles, 21 March 2022. Available from: https://www.chronicle.co.zw/why-tsholotsho-has-highest-hiv-burden/

Mabaso M, Mlangeni L, Makola L, Oladimeji O, Naidoo I, Naidoo Y, Chibi B, Zuma K, Simbayi, L, 2021. Factors Associated With Age-Disparate Sexual Partnerships Among Males And Females In South Africa: A Multinomial Analysis Of The 2012 National Population-Based Household Survey Data. Emerging Themes Epidemiol 18;3.

Makurumidze R, Decroo T, Lynen L, Chinwadzimba ZK, Van Damme W, Hakim J, Rusakaniko S, 2020. District-Level Strategies To Control The HIV Epidemic In Zimbabwe: A Practical Example Of Precision Public Health. BMC Res Notes 13;1-6.

Manepalli UR, Bham GH, Kandada S. Evaluation Of Hotspots Identification Using Kernel Density Estimation (K) And Getis-Ord (Gi*) On I-630. 3rd International Conference On Road Safety And Simulation, 2011; pp. 14-16.

Manyangadze T, Chimbari MJ, Mavhura E, 2021. Spatial Heterogeneity Association Of HIV Incidence With Socio-Economic Factors In Zimbabwe. J Geograph Res 4.

Mishra V, Assche SBV, Greener R, Vaessen M, Hong R, Ghys PD, Boerma JT, Van Assche A, Khan S, Rutstein S, 2007. HIV Infection Does Not Disproportionately Affect The Poorer In Sub-Saharan Africa. AIDS 21;S17-S28.

Mitchell, A. 1999. The ESRI Guide To Gis Analysis.

Moturi AK, Suiyanka L, Mumo E, Snow RW, Okiro EA, Macharia PM, 2022. Geographic Accessibility To Public And Private Health Facilities In Kenya In 2021: An Updated Geocoded Inventory And Spatial Analysis. Front Public Health 10;4245.

Moyo N, Maharaj P, Mambondiani L, 2017. Food Challenges Facing People Living With HIV/AIDS In Zimbabwe. Afr J AIDS Res 16;225-230.

Oseni Z, Seedat F, Kandala NB, 2018. HIV Epidemic Heterogeneity In Zimbabwe: Evidence From Successive Demographic And Health Surveys. J Biosoc Sci 50;840-852.

Oshan TM, Li Z, Kang W, Wolf LJ, Fotheringham AS, 2019. MGWR: A Python Implementation Of Multiscale Geographically Weighted Regression For Investigating Process Spatial Heterogeneity And Scale. ISPRS Int J Geo-Information 8;269.

Patel CJ, Claypool KT, Chow E, Chung M-K, Mai D, Chen J, Bendavid E. 2022. The Demographic And Socioeconomic Correlates Of Behavior And HIV Infection Status Across Sub-Saharan Africa. Comm Med 2;104.

Schaefer R, Gregson S, Takaruza A, Rhead R, Masoka T, Schur N, Anderson SJ, Nyamukapa C, 2017. Spatial Patterns Of Hiv Prevalence And Service Use In East Zimbabwe: Implications For Future Targeting Of Interventions. J Int AIDS Soc 20;21409.

StataCorp. 2017. Stata 15 Base Reference Manual. College Station, Tx: Stata Press.

Wabiri N, Shisana O, Zuma K, Freeman J, 2016. Assessing The Spatial Nonstationarity In Relationship Between Local Patterns Of HIV Infections And The Covariates In South Africa: A Geographically Weighted Regression Analysis. Spatial Spatio-Temporal Epidemiol 16;88-99.

Wheeler DC, Páez A. 2010. Geographically Weighted Regression. Handbook Of Applied Spatial Analysis. Springer.

Zimbabwe Central Statistical Office & Macro International 2007. Zimbabwe Demographic And Health Survey 2005-06. Calverton, Maryland, Usa: Central Statistical Office/Zimbabwe And Macro International.

Zimbabwe National Statistics Agency & Icf International 2016. Zimbabwe Demographic And Health Survey 2015: Final Report. Rockville, Maryland, Usa: Zimbabwe National Statistics Agency (Zimstat) And Icf International.

Zimbabwe National Statistics Agency - Zimstat & Icf International 2012. Zimbabwe Demographic And Health Survey 2010-11. Calverton, Maryland, Usa: Zimstat And Icf International.