Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/140510

Uncovering Spatial Disparities and HIV Risk Factors Among Males in KwaZulu-Natal: A Bayesian Convolutional Approach

Exaverio Chireshe[1,*], *Retius Chifurira*[1], *Jesca Mercy Batidzirai*[1], *Knowledge Chinhamu*[1], *and Ayesha B.M. Kharsany*[2,3]

[1]Statistics, School of Agriculture and Science, College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Westville Campus, Durban, South Africa
[2]Centre for the AIDS Programme of Research in South Africa (CAPRISA), University of KwaZulu-Natal, Durban, South Africa
[3]School of Laboratory Medicine & Medical Sciences, Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa

Received: 4 Apr. 2025, Revised: 18 Jun. 2025, Accepted: 19 Jul. 2025
Published online: 1 Sep. 2025

Abstract: Human immunodeficiency virus (HIV) remains a critical public health concern in South Africa, with KwaZulu-Natal (KZN) reporting the highest prevalence nationally. Despite the burden, men aged 15–49 remain underrepresented in HIV testing and treatment, exacerbating transmission risks. This study applies spatial statistical modelling to identify geographic heterogeneity and associated risk factors of HIV prevalence among this key demographic. Data were obtained from 3547 men participating in the HIV Incidence Provincial Surveillance System (HIPSS) in KZN between June 2014 and July 2015. A Bayesian convolution model was employed to capture spatially structured and unstructured random effects, using the integrated nested Laplace approximation (INLA) for efficient posterior inference. The results reveal significant spatial variation in HIV prevalence. Age emerged as a primary determinant, with the highest odds among men aged 40–44 (OR: 14.4645; 95% CI: 8.9035–23.5022). Lower educational attainment was associated with increased risk, whereas tertiary education was protective (OR: 0.6098; 95% CI: 0.3733–0.9866). Additional covariates positively associated with HIV infection included drug use, lack of circumcision, TB and STI diagnoses, mobility, and reported use of PrEP. A spatial cluster near Pietermaritzburg, with a 0.45 km radius, exhibited a 55.6% prevalence and a relative risk of 1.66, although this was not statistically significant ($p = 0.213$). In conclusion, the study highlights the utility of Bayesian spatial models in identifying localised risk patterns and informs recommendations for geographically stratified, male-specific interventions, including integrated service delivery for HIV, TB, and STIs within high-prevalence clusters.

Keywords: Bayesian convolutional model, HIV prevalence, Kulldorff's spatial scan statistic, odds ratios, spatial clustering.

1 Introduction

Human immunodeficiency virus (HIV) remains a leading public health challenge, disproportionately affecting sub-Saharan Africa, which accounts for nearly 70% of global cases [1]. South Africa, particularly KwaZulu-Natal (KZN), is the epicentre of the epidemic, reporting some of the highest HIV prevalence rates globally [2]. While significant strides have been made in expanding access to antiretroviral therapy (ART) and reducing new infections, gender disparities persist. Men remain significantly underrepresented in testing, treatment, and care programmes, contributing to higher mortality rates and ongoing transmission.

The epidemic among men is particularly concerning as they are less likely than women to access HIV testing and care services [3,4,5], achieve viral suppression due to delayed initiation of ART [6,7], and participate in prevention programmes such as voluntary medical male circumcision [8]. Despite the recognised importance of involving men in HIV prevention strategies, most intervention programmes have historically targeted women and children, driven by the maternal and child health agenda. This has created a gap in understanding male-specific risk factors and designing interventions tailored to this population [9].

* Corresponding author e-mail: ekichireshe@gmail.com

Incorporating spatial modelling into the analysis of HIV prevalence allows for an understanding of geographic heterogeneities, highlighting hotspots and underlying risk factors unique to men aged 15–49. This age group is pivotal because it represents sexually active males, who play a critical role in HIV transmission dynamics and bear the brunt of socio-structural factors influencing their vulnerability [10].

Research highlights significant spatial variations in HIV prevalence across regions in South Africa, emphasising the need for geographically tailored strategies. Work by [11] identified substantial disparities in HIV prevalence and called for targeted interventions informed by spatial analyses. Advanced methods, such as Bayesian spatial models, have demonstrated the ability to capture these variations effectively. A study by [12] employed Bayesian spatial models to identify key predictors and spatial clusters of HIV prevalence among females aged 15–34 in uMgungundlovu District, KZN, South Africa. While this study focused on females, similar methodologies can be applied to understand risk factors among males.

Geographical hotspots of HIV prevalence are often linked to socioeconomic inequalities and limited access to healthcare [13]. Studies have demonstrated that traditional statistical methods fail to adequately account for spatial dependencies and heterogeneity, underscoring the potential of advanced Bayesian methods, such as convolution models, to address these gaps. For example, [14] utilised Bayesian hierarchical spatial modelling to explore the clustering of HIV prevalence at the district level, while [15] applied spatial analysis to identify underserved areas with high HIV prevalence, advocating for geographically specific strategies.

Shifts in HIV incidence patterns across age groups have also been observed in KwaZulu-Natal. Work by [16] highlighted the need for age-specific interventions, particularly among males aged 15–49, who contribute significantly to epidemic dynamics [17]. Despite this, much of the existing research has failed to disaggregate data by age or gender, limiting insights into age-related risks among males or spatial variability in prevalence.

The role of socioeconomic factors, such as education and employment, in influencing HIV prevalence is well-documented [18,19]. However, little is known about how these determinants vary across spatial contexts in KwaZulu-Natal. Behavioural risk factors, including mobility, substance abuse, and male circumcision, have often been studied in isolation. Their combined impact and interaction with spatial disparities remain poorly understood, limiting the effectiveness of current intervention strategies.

To date, limited research has integrated individual, behavioural, and spatial determinants of HIV prevalence among males aged 15–49 in KwaZulu-Natal. Bridging this gap is crucial for designing interventions that address both individual-level risk factors and geographic disparities. This study seeks to address these gaps by applying a Bayesian convolution model to explore the spatial disparities and key risk factors associated with HIV prevalence among males aged 15–49 in KwaZulu-Natal. By incorporating spatial heterogeneity, this analysis provides critical insights into localised risks and offers evidence for spatially informed policy and intervention strategies.

2 Materials and Methods

2.1 Study Area Location

Figures 1A and 1B below depict the location of the study area within uMgungundlovu District and the location of the two sub-districts of Vulindlela (Western part) and the Greater Edendale (Eastern part), respectively.

2.2 Sources of Data and Study Population

This study used secondary data from the HIV Incidence Provincial Surveillance System (HIPSS), a population-based survey conducted from June 2014 to July 2015 in Vulindlela (rural) and Greater Edendale (peri-urban) areas of uMgungundlovu District, KwaZulu-Natal, South Africa. HIPSS is known for its robust design in assessing HIV incidence and prevalence. A multi-stage probability sampling method was employed. Out of the 600 enumeration areas (EAs), 221 were randomly selected from 591 eligible EAs (each with over 50 households). Within selected EAs, households were chosen systematically, and one eligible individual per household was randomly selected after providing written informed consent. GPS coordinates were recorded for all sampled households to ensure spatial precision and reduce selection bias. To maintain data quality, the Mobenzi Researcher system enabled real-time monitoring of field teams. Data quality was checked daily during the first month, monthly for six months, and quarterly thereafter. Automated quality control tools flagged inconsistencies for immediate correction. Laboratory-confirmed HIV test results from peripheral blood samples were also integrated. Through rigorous sampling, real-time monitoring, and multi-tiered quality control, HIPSS ensured a reliable and representative dataset of males aged 15–49. These measures underpin the study's credibility in analysing demographic, behavioural, and socioeconomic risk factors associated with HIV

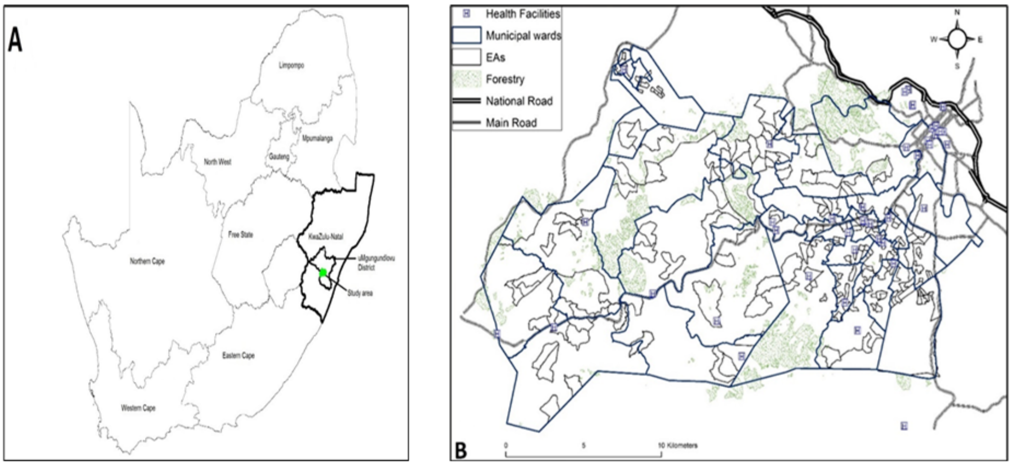

Fig. 1: 1A and 1B. Location of the study area.

prevalence. The HIPSS survey enrolled 9,812 participants aged 15–49, including 6,265 females and 3,547 males. For this study, 2,315 males were analysed after excluding cases with missing HIV status or key demographic data. A complete-case analysis was applied, excluding participants with incomplete records. While this method may introduce selection bias if excluded individuals differ systematically, it ensures analytical consistency, and the impact on conclusions is likely minimal. The study targeted males aged 15–49, a group central to HIV transmission due to their sexually active and economically productive age. Understanding the spatial disparities and risk factors within this demographic is crucial for designing targeted, evidence-based interventions.

2.3 Study Variables

The dependent variable was HIV prevalence, defined as the proportion of HIV-positive participants in an enumeration area (EA). Unweighted prevalence was used to better identify geographic clustering patterns.

HIV status among participants in the study population was categorised as a binary outcome:

$$y_{ij} = \begin{cases} 1 & \text{for a participant who tested positive} \\ 0 & \text{for a participant who tested negative} \end{cases} \tag{1}$$

Sociodemographic, behavioural, and biological covariates were selected based on epidemiological relevance, data availability, and statistical significance. Variables identified in previous research as key HIV risk factors were prioritised, and only those with minimal missing data were retained to reduce bias. Univariate analyses informed the selection of significant predictors for the multivariate model. The final set of covariates included age, level of education, marital status, main income, ever consumed alcohol, ever diagnosed with tuberculosis (TB) or a sexually transmitted infection (STI), number of sexual partners, condom use, male circumcision status, drug abuse, access to healthcare, Pre-Exposure Prophylaxis (PrEP), time spent away from home, and length of residence in the community. Age was categorised into seven groups: 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, and 45–49, to reflect distinct behavioural and socio-biological stages. This granularity supports targeted analysis and intervention design. Multicollinearity was assessed using the variance inflation factor (VIF), with all values below 1.5, indicating minimal collinearity. A stepwise selection process excluded non-significant or redundant variables, yielding a more streamlined and interpretable model. Spatial dependency was explicitly modelled to account for unmeasured geographic factors, reducing omitted variable bias and enhancing accuracy.

2.4 Spatial Autocorrelation

The Bayesian convolution regression model was chosen for its strength in capturing spatial dependencies, handling heterogeneity, and managing uncertainty in HIV prevalence among males aged 15–49 in KwaZulu-Natal. Unlike frequentist models, which assume independence, the Bayesian approach incorporates spatial autocorrelation, improving geographic risk estimation. Given the spatial clustering observed, this model was well-suited. Spatial analysis used geo-referenced Enumeration Areas (EAs) linked to HIV data. Since nearby areas often show similar values, spatial models help distinguish true geographic trends from random variation. Moran's I and Geary's C confirmed spatial structure in the data.

2.4.1 Global Moran's Index Statistic

The global Moran's index statistic assesses the overall spatial autocorrelation across the whole study area. The advantage of this statistic is its capability to examine the existence, tendency and potency of spatial autocorrelation over a given area. Positive spatial autocorrelation is depicted when values in the neighbouring enumeration areas are comparable or are spatially grouped [21]. On the other hand, when adjacent enumeration areas tend to have contrasting values or are spatially scattered, Moran's index statistic suggests that there is a negative spatial autocorrelation [22,23], and when some neighbouring pair's deviations are in the same direction and others in the opposite direction, it approaches zero [24].

The formula for calculating Moran's index statistic is given by:

$$I = \frac{n}{\Theta} \cdot \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \Theta_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2}$$

where n is the total number of enumeration areas, y_i is the value of the variable at location i, y_j is the value of the variable at a neighbouring location j, $\bar{y}$ is the mean of the variable y across all enumeration areas, Θ_{ij} is the spatial weight between enumeration area i and enumeration area j, and Θ indicates the sum of all spatial weights.

2.4.2 Geary's C Statistic

Geary's C statistic measures the spatial autocorrelation by evaluating the degree of similarity or dissimilarity between values of a variable at neighbouring spatial locations. Geary's C statistic is more sensitive to local spatial variations than Moran's index statistic, which assesses global spatial patterns [25,26]. The formula for Geary's C is expressed as follows:

$$C = \frac{(N-1)\sum_{i=1}^{N} \sum_{j=1}^{N} \Theta_{ij}(y_i - y_j)^2}{2\Theta \sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{3}$$

where N is the total number of enumeration areas (locations), y_i and y_j are the values of the variable of interest at locations i and j, $\bar{y}$ is the mean of the variable across all locations, Θ_{ij} is the spatial weight between location i and location j, and Θ is the sum of all Θ_{ij}. The value of C lies between 0 and 2. If C = 1, there is no spatial autocorrelation; if $C < 1$, there is positive spatial autocorrelation; and negative spatial autocorrelation is represented by values of $C > 1$ [27].

To further identify and interpret spatial clusters, we utilised Kulldorff's spatial scan statistic (SaTScan). Unlike Moran's I and Geary's C, which examine general spatial patterns, SaTScan is designed to detect significant clusters of high-risk (hot-spots) and low-risk (cold-spots) areas by scanning circular windows across the study region. This method enables the precise identification of areas with notably higher or lower HIV prevalence, providing valuable insights for targeted public health interventions [28].

2.5 Bayesian Hierarchical Spatial Models

Bayesian hierarchical modelling is a multi-level statistical approach that estimates posterior distributions using Bayesian methods. It is well-suited for analysing areal data, where outcomes are aggregated by geographic units [29]. These models incorporate covariates and random effects to account for unexplained variation and spatial autocorrelation, offering flexibility in quantifying uncertainty and assessing covariate effects. By borrowing information from neighbouring areas, they produce smoothed relative risk estimates [30]. In this study, we apply a spatial Bayesian smoothing logistic regression model.

One of the key properties of spatial data is *spatial dependence*. In this study, this property was considered as follows: a participant in a ward, say ε_j, of the uMgungundlovu District carries some information about HIV that is also observed in wards that are geographically close to ε_j.

Suppose that the outcome variable y_{ij} represents the HIV test result for the ith participant in the ε_jth ward, with associated probability p_{ij}, where:

$$y_{ij} = \begin{cases} 1 & \text{if the participant tested positive,} \\ 0 & \text{if the participant tested negative,} \end{cases} \quad i = 1,\ldots,n_j; \quad j = 1,\ldots,221,$$

and n_j is the number of participants in the ε_j ward included in the 2014 HIPSS baseline survey. Our outcome variable had a binary outcome, and hence, it is necessary to fit a Bayesian hierarchical spatial smoothing logistic regression model. The model is given by:

$$\omega_{ij} = \beta(p_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\alpha}_i + \mu_j + \phi_j, \quad i = 1,\ldots,n_j; \quad j = 1,\ldots,221, \tag{4}$$

where $\beta(\cdot)$ is the link function, $\mathbf{x}_{ij} = (1, x_{1ij}, \ldots, x_{pij})$ is the vector of p covariates measured on the ith participant in the ε_j ward, and $\boldsymbol{\alpha}_i$ is the vector of regression coefficients, including the intercept.

The term μ_j is an unstructured random effect assumed to follow a normal distribution, $\mu_j \sim \mathcal{N}(0, \sigma^2_\mu)$, and ϕ_j represents the structured spatial random effect for ward ε_j.

2.5.1 Bayesian Parameter Estimation Framework

In the Bayesian framework, parameter estimation is not simply about finding point estimates (e.g., maximum likelihood estimates), but about characterising the full posterior distribution of the parameters. This allows us to quantify uncertainty and incorporate prior beliefs into the analysis [31,32,33].

Bayesian estimation is based on *Bayes' theorem*, which provides a formal way to update our beliefs about unknown parameters based on observed data. It is given by:

$$P(\boldsymbol{\beta} \mid \mathbf{y}) = \frac{P(\mathbf{y} \mid \boldsymbol{\beta})\, P(\boldsymbol{\beta})}{P(\mathbf{y})} \tag{5}$$

where $P(\boldsymbol{\beta} \mid \mathbf{y})$ is the *posterior distribution* of the parameter $\boldsymbol{\beta}$ after observing the data $\mathbf{y}$; $P(\mathbf{y} \mid \boldsymbol{\beta})$ is the *likelihood function*, which describes how likely the observed data are given the parameter values $\boldsymbol{\beta}$; $P(\boldsymbol{\beta})$ is the *prior distribution*, representing prior beliefs about the parameter before seeing the data; and $P(\mathbf{y})$ is the *marginal likelihood*, serving as a normalising constant.

The marginal likelihood distribution normalises the posterior distribution and ensures that the posterior is a valid probability distribution and sums to 1. It is obtained by integrating the joint distribution over all possible values of β and is given by:

$$P(\mathbf{y}) = \int P(\mathbf{y} \mid \boldsymbol{\beta})\, P(\boldsymbol{\beta})\, d\boldsymbol{\beta}. \tag{6}$$

2.5.2 Prior Distributions

In the Bayesian framework, prior distributions (representing beliefs before observing data) are combined with likelihood functions (representing observed data) to yield posterior distributions (updated beliefs). Priors are essential in Bayesian hierarchical spatial models, particularly for managing small sample sizes or high spatial variability [31,34,35,36]. Priors must be specified for all random components. When prior knowledge is strong, informative priors are used; otherwise, non-informative or weakly informative priors are preferred. In this study, we used non-informative priors for regression coefficients and random effect variances, given the limited prior knowledge available. We adopted penalised complexity (PC) priors for the precision of random effects, as they strike a balance between model simplicity and complexity, allowing the data to inform the level of spatial structure. Unlike flat or overly vague priors, PC priors reduce the risks of overfitting and improve interpretability and computation [37,38].

The PC prior for precision ρ is derived to penalise deviation from a simpler base model. For precision, the PC prior is expressed as:

$$\pi(\rho) = \nu e^{-\nu\sqrt{\rho}} \tag{7}$$

with

$$\nu = \frac{\log(\alpha)}{U} \tag{8}$$

and

$$\rho = \frac{1}{\sigma^2} \tag{9}$$

where ρ is the precision, U is the upper bound for the standard deviation σ of the random effect, and α is the probability that $\sigma > U$.

2.5.3 Posterior Distributions and Point Estimates

The complete information about the parameter estimates is contained in the posterior distribution, and point estimates and credible intervals are generally used to summarise the posterior distribution. Point estimates are divided into three categories, namely posterior mean, posterior mode and posterior median. The posterior distribution is then used for inference and prediction.

The posterior mean is the expected value of the parameter under the posterior distribution. It is a common estimate, especially when the posterior is symmetric. For a parameter β, it is given by:

$$\hat{\boldsymbol{\beta}}_{\text{mean}} = \mathbb{E}[\boldsymbol{\beta} \mid \mathbf{y}] = \int \boldsymbol{\beta}\, P(\boldsymbol{\beta} \mid \mathbf{y})\, d\boldsymbol{\beta} \tag{10}$$

The posterior mode is a point estimate of a parameter in the Bayesian framework. It corresponds to the value of β that maximises the posterior distribution $P(\beta \mid y)$, that is, the most probable value of β given the observed data. It is expressed mathematically as:

$$\hat{\boldsymbol{\beta}}_{\text{MAP}} = \arg\max_{\boldsymbol{\beta}} P(\boldsymbol{\beta} \mid \mathbf{y}) \tag{11}$$

where $\arg\max_{\beta}$ indicates finding the value of β that maximises this posterior probability. The MAP estimate is often used when the posterior is skewed, but it can be sensitive to the choice of the prior.

The posterior median is a robust point estimate that divides the posterior distribution into two equal parts. It is less sensitive to outliers compared to the mean or mode.

Bayesian credible intervals indicate the range within which the parameter is likely to fall with a certain probability. A 95% credible interval means that there is a 95% probability that the true parameter lies within the interval, given the data and prior.

For a parameter $\boldsymbol{\beta}$, the 95% credible interval is given by:

$$P(\boldsymbol{\beta}_{\text{lower}} < \boldsymbol{\beta} < \boldsymbol{\beta}_{\text{upper}} \mid \mathbf{y}) = 0.95. \tag{12}$$

Unlike frequentist confidence intervals, credible intervals directly provide the probability that the parameter lies within a certain range.

2.5.4 Parameter Estimation Computational Techniques

In most Bayesian practical problems having models with hierarchical or spatial structures, the posterior distribution does not have a closed-form solution. Hence, we depend on computational techniques to approximate the posterior distribution. These computational techniques include Markov Chain Monte Carlo (MCMC), Integrated Nested Laplace Approximation (INLA) and Variational Inference (VI).

MCMC techniques are utilised to estimate the posterior distribution by creating samples from it. These samples can then be utilised to approximate posterior summaries such as means, medians, and credible intervals. Gibbs Sampling, Metropolis-Hastings Algorithm, and Hamiltonian Monte Carlo (HMC) are the most widely applied MCMC techniques [39,40,41].

The Gibbs sampling technique involves sampling the posterior, one parameter at a time, from its full conditional distribution and iteratively updating each parameter. It can be expressed as:

$$\boldsymbol{\beta}_i^{(t+1)} \sim P(\beta_i \mid \boldsymbol{\beta}_{-i}, \mathbf{y})$$

The Metropolis–Hastings algorithm is a more general MCMC method applied when the full conditionals are not easily sampled. A candidate point is proposed at each iteration, and it is accepted or rejected based on the acceptance ratio given by:

$$\alpha = \min\left(1, \frac{P(\boldsymbol{\beta}^* \mid \mathbf{y})\, q(\boldsymbol{\beta}^{(t)} \mid \boldsymbol{\beta}^*)}{P(\boldsymbol{\beta}^{(t)} \mid \mathbf{y})\, q(\boldsymbol{\beta}^* \mid \boldsymbol{\beta}^{(t)})}\right), \tag{13}$$

where $\boldsymbol{\beta}^*$ is the proposed new parameter value, and $q(\boldsymbol{\beta} \mid \boldsymbol{\beta}')$ is the proposal distribution.

The Hamiltonian Monte Carlo (HMC) is a more advanced MCMC algorithm that utilises gradients of the log-posterior to examine the parameter space more conveniently, resulting in faster convergence. An alternative to MCMC is INLA, which is designed specifically for latent Gaussian models. This computational technique uses numerical approximations instead of sampling to compute the posterior distribution. It is suitable for large sets of data or complex hierarchical models and is faster and provides accurate posterior marginal distributions for each parameter compared to MCMC [42,43,44].

The Variational Inference (VI) is a method that estimates the posterior by getting a simpler distribution that is close to the true posterior. Instead of sampling, it minimises the Kullback-Leibler (KL) divergence between the true posterior and the approximating distribution [45,46,47]
.

In our analysis, we preferred INLA due to its speed, accuracy, scalability, and ease of use. It outperforms MCMC in these contexts by eliminating the need for sampling while delivering precise results [42].

2.6 Bayesian Convolution Model

The Bayesian Convolution Model provides a robust, data-driven approach to analysing HIV prevalence, accounting for spatial structure, uncertainty, and small-area variation. This makes it an ideal choice for uncovering spatial disparities and risk factors among males aged 15-49 in KZN, ultimately helping to inform more effective HIV intervention strategies.

In convolution models, both unstructured and structured components are combined to model both spatially structured effects and unstructured heterogeneity simultaneously. In our study, the Bayesian convolution model deals with a binary outcome variable Y_i for each observation i, where $Y_i \in \{0,1\}$ and follows a Bernoulli distribution. Thus,

$$Y_i \sim \text{Bernoulli}(p_i),$$

where p_i is the probability that $Y_i = 1$.

The link function is the logistic function given by:

$$p_i = \frac{\exp(\phi_i)}{1+\exp(\phi_i)}, \tag{14}$$

where ϕ_i is a linear predictor expressed as

$$\phi_i = \beta_0 + \mathbf{X}_i^T \boldsymbol{\beta}, \tag{15}$$

where β_0 is the intercept, $\mathbf{X}_i$ is the vector of covariates for observation i, and $\boldsymbol{\beta}$ is the vector of regression coefficients.

2.6.1 Convolution Model Specification

To account for spatial autocorrelation and unstructured heterogeneity, the convolution model extends the logistic regression formulation by incorporating both structured and unstructured random effects:

$$\text{logit}(p_i) = \beta_0 + \mathbf{X}_i^T \boldsymbol{\beta} + u_i + \theta_i \tag{16}$$

where u_i denotes the unstructured random effects and $u_i \sim \mathcal{N}(0, \sigma_u^2)$, and θ_i indicates the spatially structured random effect at location i, typically modelled using an intrinsic conditional autoregressive (ICAR) prior or a Besag–York–Mollié (BYM) model to capture spatial dependence.

2.6.2 Likelihood Function

Given the Bernoulli assumption for Y_i, the likelihood function for the Bayesian convolution logistic model is:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} p_i^{Y_i} (1-p_i)^{1-Y_i} \tag{17}$$

Substituting the logit function:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} \left(\frac{\exp(\beta_0 + \mathbf{X}_i^T \boldsymbol{\beta} + u_i + \theta_i)}{1+\exp(\beta_0 + \mathbf{X}_i^T \boldsymbol{\beta} + u_i + \theta_i)} \right)^{Y_i} \left(\frac{1}{1+\exp(\beta_0 + \mathbf{X}_i^T \boldsymbol{\beta} + u_i + \theta_i)} \right)^{1-Y_i} \tag{18}$$

2.7 Model Diagnostics

After fitting the model, we checked for the model adequacy using the residuals plot and the normal Q-Q plot. If the residuals plot shows that residuals are not symmetrically distributed about zero, or are showing any clear pattern or trend, and have no constant variance, then the model will not be of best fit [48,49,50]. The deviation from normality in the Q-Q plot suggests that the residuals do not perfectly follow a normal distribution.

We also examined the spatial autocorrelation in the residuals by applying Moran's index statistic, Geary's C statistic, and the variogram plot. We applied these to verify whether the spatial structure was adequately captured. The model's failure to fully capture spatial dependencies is indicated by high spatial autocorrelation in the residuals [51,52]. If Moran's index statistic and Geary's C statistic for the residuals are significant, then the model will not fit the data well. Using the variogram plot, an increasing semi-variance with distance implies spatial autocorrelation, suggesting that the model may not have fully accounted for the spatial structure. On the other hand, a flat variogram implies that residuals are spatially uncorrelated, indicating that the model has adequately captured the spatial structure [51,53].

Posterior density plots were also used to examine the validity, reliability and stability of the fitted model. A smooth, unimodal plot suggests the model has fit well and provides reliable results for the parameter under consideration. Multimodal may suggest ambiguity or issues with the model or data [31].

2.8 Software and Implementation

The implementation of the Bayesian convolution model was carried out using the integrated nested Laplace approximation (INLA) method, as developed by [42] and later reviewed in 2017. This was performed using R software (version 4.4.0) with the application of the “INLA”, “sf”, “sp”, “spdep”, and “dplyr” packages. Spatial relationships between the enumeration areas were established by creating a spatial weight matrix, with neighbours identified through Queen’s contiguity. We also employed Kulldorff’s spatial scan statistics via SaTScan software version 10.1.3.

3 Results, Discussion and Conclusion

3.1 Results

3.1.1 Descriptive Statistics Results

Table 1 presents the HIV prevalence (%) and corresponding 95% confidence intervals for all variables included in the study. A total of 2,315 males were included in the analysis after excluding participants with incomplete HIV status or missing key demographic variables. Among them, 787 were HIV-positive, resulting in an overall HIV prevalence of 34.0% (95% CI: 32.07–35.93).

HIV prevalence showed clear variation across demographic and behavioural factors. Prevalence increased with age, peaking at 62.5% in the 40–44 age group, while the lowest prevalence was observed in the 15–19 age group (4.55%), highlighting age as a critical risk factor. Educational attainment was inversely associated with HIV prevalence; males with tertiary education had the lowest prevalence (16.13%), whereas those with only primary education or no schooling had significantly higher rates. Unemployed males and those relying on remittances also exhibited differing prevalence levels, with the highest (40.31%) observed among those with no income. ¿p5cm r c c c c

Covariate	n = 787	Prevalence (%)	95% CI Lower	95% CI Upper	p-Value

Covariate n = 787 Prevalence (%) 95% CI Lower 95% CI Upper p-Value

Age Group					< 0.0001
15–19	10	4.55	1.82	7.27	
20–24	61	11.01	8.48	13.72	
25–29	140	30.77	26.59	34.95	
30–34	184	50.00	44.84	55.16	
35–39	160	54.42	48.64	60.20	
40–44	145	62.50	56.03	68.53	
45–49	87	45.31	38.54	52.60	
Education Level					< 0.0001
Complete Secondary	285	27.67	24.95	30.39	
Incomplete Secondary	396	41.25	38.12	44.38	
No Schooling/Pre-primary	20	43.48	30.43	58.70	
Primary (Grade 1–7)	61	49.19	40.32	58.06	
Tertiary	25	16.13	10.32	21.94	
Main Income					0.0237
No Income	79	40.31	33.67	47.45	
No Response	18	36.00	24.00	50.00	
Other non-farming income	87	36.71	30.80	43.04	
Pension or grants	150	33.86	29.57	38.37	
Remittance	9	23.68	10.53	36.84	
Salary/Wage	441	32.86	30.33	35.39	
Farming sales	3	33.33	0.00	66.67	
Marital Status					0.1732
Divorced	0	0.00	0.00	0.00	
Legally married	66	40.49	33.13	47.85	

Living together 37 64.91 52.63 77.19
Separated 1 50.00 0.00 100.00
Single (never married) 623 31.50 29.47 33.57
Single (previously cohabiting) 58 52.73 43.64 61.82
Widowed 2 66.67 0.00 100.00

Drug Abuse 0.5179
Never 762 34.37 32.39 36.36
Yes 25 25.51 17.35 34.69

PrEP Use 0.5272
No 778 33.78 31.87 35.74
Yes 9 75.00 50.00 100.00

Circumcision Status < 0.0001
Circumcised 161 20.05 17.31 22.91
Uncircumcised 626 41.40 38.96 43.92

Ever Diagnosed with TB 0.4287
No 723 33.20 31.22 35.17
No Response 2 20.00 0.00 50.00
Yes 62 48.82 40.16 57.48

Condom Use 0.2837
No 26 37.14 25.71 48.57
Yes 761 33.90 31.94 35.86

Number of Sexual Partners 0.7022
1 598 34.59 32.33 36.84
2 75 28.09 22.85 33.71
3+ 114 35.74 30.41 41.07

Alcohol Use 0.8114
No 383 32.82 30.16 35.56
Yes 404 35.19 32.40 37.98

Diagnosed with STI 0.0533
No 694 33.02 31.02 35.01
Yes 93 43.66 37.09 50.23

Away from Home 0.0035
No 697 34.22 32.16 36.28
No response 0 0.00 0.00 0.00
Yes 90 33.46 27.88 39.03

Length in Community 0.0094
Always 586 32.81 30.63 34.99
<1 year 17 33.33 21.57 47.06
>1 year 183 38.45 34.03 42.86
No response 1 50.00 0.00 100.00

Access to Healthcare 0.3380
Did not respond 1 14.29 0.00 42.86
No 515 33.03 30.72 35.41
Yes 271 36.18 32.71 39.65

Marital status, health behaviour, and biomedical factors further influenced HIV prevalence. Widowers (66.67%) and cohabiting males (64.91%) had the highest rates. Surprisingly, PrEP users had a much higher prevalence (75%), likely due to reverse causality (PrEP targeting high-risk individuals). Uncircumcised males, those diagnosed with TB or STIs, and those reporting multiple sexual partners or alcohol consumption all had higher prevalence. Interestingly, some factors such as condom use, time spent away from home, recent migration, and access to healthcare showed only marginal differences, with slightly higher prevalence among males with access to services, potentially reflecting greater health-seeking behaviour in higher-risk populations.

3.1.2 Spatial Variation and Clustering of HIV Prevalence

HIV prevalence varied across geographical locations, as illustrated in Figure 2. The spatial distribution map, based on posterior estimates from the Bayesian convolution model, depicts the geographic variation in HIV prevalence using posterior mean estimates. The colour gradient represents these estimates, highlighting spatial heterogeneity in HIV prevalence across different regions. The estimated posterior means ranged from -0.10 to 0.15, indicating that HIV prevalence is not uniformly distributed across the study area. The global Moran's index statistic displayed in Table 2, was

Fig. 2: Spatial Distribution Map of HIV prevalence (Interactive Map), see: https://drive.google.com/file/d/1UyV2lMSczfmsHpaww2qnkQoDh18bTWup/view?usp=sharing

0.7699 with a p-value < 0.001, indicating the existence of significant positive spatial autocorrelation in the uMgungundlovu District. Since the Moran's index value is positive and statistically significant, we can conclude that HIV prevalence among males in the district is spatially heterogeneous. This suggests the presence of clusters of high and low HIV prevalence areas within the study region, indicating a non-random spatial pattern. The positive Moran's index also implies that HIV prevalence in neighbouring wards tends to be similar.

These findings are supported by the results of Geary's C statistic, which further confirm the presence of spatial heterogeneity in HIV prevalence among males within the uMgungundlovu District. The summary statistics for Moran's index and Geary's C are provided in Table 2.

Table 1: Moran's I and Geary's C Summary Statistics

Summary Statistics	Moran's I	Geary's C
Statistic	0.7699	0.2180
p-value	< 0.001	< 0.001
Expectation	-0.0004	1.0000
Variance	0.0002	0.0002
Standard Deviate	62.3053	53.0581

The results from Moran's Index and Geary's C consistently reveal spatial heterogeneity in HIV prevalence among males in the uMgungundlovu District. These findings imply that HIV prevalence is not randomly distributed across the district, but rather spatially clustered.

However, Moran's index test statistic and Geary's C statistic are less informative when coming to distinguishing between hotspots and cold spots clusters. Accordingly, to identify hotspots and cold-spots wards for HIV prevalence, we employed Kulldorf's spatial scan statistic. This statistic identified one cluster. The distribution of spatial clusters of HIV prevalence is displayed on a map in Figure 3 below.

Fig. 3: Spatial Clustering of HIV Prevalence in uMgungundlovu Municipality, see: https://drive.google.com/file/d/1xMOxYpbpMQA9vjd38d9UpuvD5DVU_mPM/view?usp=sharing

The identified cluster, which was a hotspot with a radius of 0.45 km, had a 55.6% HIV prevalence and a relative risk of 1.66. Thus, the risk within the cluster was 66% higher than outside the cluster. This cluster was around the Pietermaritzburg area, with five location IDs. However, this cluster was not statistically significant since its p-value was $0.213 > 0.05$.

3.1.3 Bayesian Spatial Modelling and Risk Factor Analysis

Summary statistics for HIV prevalence rates across all covariates included in the study are presented in Table 1. While these statistics provide an initial indication of potential associations, the convolution model results are emphasised due to their robustness in accounting for spatial correlations and confounding effects. This approach ensures that conclusions are drawn from a more comprehensive analysis of the data.

To investigate spatial disparities and risk factors associated with HIV prevalence among males aged 15–49 in KwaZulu-Natal, a Bayesian convolution model was fitted. Most covariates were statistically significant at the 5% significance level, though a few did not demonstrate statistical significance.

To identify factors associated with HIV prevalence, adjusted odds ratios (ORs) were calculated using the fitted convolution model, implemented with INLA in R software. Table 3 presents the adjusted ORs along with their corresponding 95% credible intervals (CIs) for the participants' characteristics. ¿p6.2cmccc

Covariate OR 95% CI Lower 95% CI Upper

Covariate	**OR**	**95% CI Lower**	**95% CI Upper**
Intercept	0.0488	0.0098	0.2430
Age Group (ref: 15–19)			
20–24	1.4234	0.8834	2.2934
25–29	4.5357	2.8950	7.1068
30–34	9.9118	6.3024	15.5902
35–39	11.7944	7.4068	18.7837
40–44	14.4645	8.9035	23.5022
45–49	7.0116	4.2016	11.7017
Education (ref: Complete Secondary)			
Incomplete secondary (Grade 8–11/NTC1/2)	1.6192	1.3061	2.0075
No schooling/creche/pre-primary	1.4539	0.7685	2.7499

Primary (Grade 1–7) 1.6389 1.0751 2.4990
Tertiary (Diploma/degree) 0.6098 0.3733 0.9886

Main Income (ref: No Income)
No response 1.0677 0.5364 2.1253
Other non-farming income 0.9185 0.5996 1.4068
Pension or grants 0.8064 0.5501 1.1823
Remittance (migrant worker) 0.5863 0.2568 1.3384
Salary and/or wage 0.7901 0.5666 1.1017
Sales of farming products 0.7926 0.2303 2.7280

Marital Status (ref: Divorced)
Legally married 0.6970 0.2795 1.7382
Living together (like husband and wife) 1.7158 0.6441 4.5704
Separated, still legally married 0.8891 0.1729 4.5728
Single, never married 0.9868 0.4077 2.3880
Single, lived with someone before 1.3297 0.5236 3.3769
Widowed 1.6124 0.3416 7.6106

Drug Abuse (ref: Never)
Yes 1.6226 1.0357 2.6600

PrEP (ref: No)
Yes 3.1180 1.0001 9.7225

Male Circumcision Status (ref: Circumcised)
Uncircumcised 2.0533 1.6424 2.5673

TB (ref: No)
No response 0.6010 0.1769 2.0416
Yes 1.9868 1.3087 3.0165

Condom Use (ref: No)
Yes 0.8183 0.4705 1.4232

Number of Partners (ref: 1)
2 0.8021 0.5769 1.1152
3 1.0882 0.8085 1.4642

Alcohol (ref: No)
Yes 0.9701 0.7958 1.1828

STI Diagnosed (ref: No)
Yes 1.4468 1.0448 2.0042

Away From Home (ref: No)
No response 0.2850 0.0684 1.1873
Yes 1.3163 1.0772 1.6400

Length in Community (ref: Always)
Moved here less than 1 year ago 0.9863 0.5257 1.8497
Moved here more than 1 year ago 1.0597 0.8306 1.3518
No response 1.6147 0.2961 8.8049

Accessed Health Care (ref: No response)
No 1.2765 0.4085 3.9893
Yes 1.6223 0.5166 5.0953

The covariates depicted in Table 3 provide insights into the factors associated with HIV prevalence among males aged 15-49 in KwaZulu Natal. Only significant covariates are reported here. A covariate is significant if its 95% credible interval does not include 1.

Age emerged as a strong determinant of HIV prevalence, with older males, especially those aged 30–44, at significantly higher risk. Men aged 25–29, 30–34, 35–39, 40–44, and 45–49 showed significantly higher odds of being HIV-positive compared to those aged 15–19. The odds ratios for these age groups were 4.5357, 9.118, 11.7944, 14.4645 and 7.0116 respectively. The risk peaked in the 40–44 age group (OR: 14.4645; 95% CI: 8.9035–23.5022), indicating that this group has 14.4645 times higher odds of HIV prevalence compared to the age group 15-19.

Considering the level of education, the odds of HIV prevalence for males with incomplete secondary were 1.6192 (OR = 1.6192, 95% CI: 1.3061 – 2.0075) times more than those with complete secondary. Participants with primary education were 1.6389 (OR = 1.6389, 95% CI: 1.0751 – 2.4990) times more likely to be HIV infected compared to participants with complete secondary. Males with tertiary education were at lower risk (OR = 0.6098, 95% CI: 0.3733 – 0.9866) of being HIV infected compared to those with complete secondary, suggesting that higher education may be protective.

Based on the drug abuse covariate, males who reported drug abuse had higher odds of being HIV-positive (OR: 1.6226; 95% CI: 1.0357–2.6600). For the PREP covariate, males who reported using PrEP had an increased risk of being HIV-infected (OR = 3.1180, 95% CI: 1.0001 – 9.7225). Male circumcision also emerged as a key factor for HIV prevalence, with uncircumcised males having significantly higher odds of being HIV-positive (OR: 2.0533; 95% CI: 1.6424–2.5673) compared to circumcised males. Males who were diagnosed with TB were 1.9868 (OR =1.9868, 95% CI: 1.3087 – 3.0165) times more likely to be HIV infected compared to those who never suffered from TB. The odds of HIV prevalence for males who were diagnosed with STIs were 1.4468 times greater than those who were not diagnosed with STIs.

Results based on being away from home revealed that males who were away from home showed higher odds of HIV prevalence (OR: 1.3163; 95% CI: 1.0772–1.6400) compared to those who have not been away from home. This suggests that mobility and migration may increase exposure to high-risk environments.

From the interpretations above, the results highlight significant spatial disparities and risk factors associated with HIV prevalence among males in KwaZulu-Natal. The key significant risk factors identified include older age, lower educational attainment, lack of circumcision, drug abuse, TB, STIs, and mobility (being away from home).

3.1.4 Diagnostic Evaluation of the Convolution Model

The performance of the fitted convolution model was evaluated using residual plots and normal quantile-quantile (Q-Q) plots. To assess spatial autocorrelation in the residuals, Moran's Index, Geary's C statistic, and the variogram plot were applied. Additionally, posterior density plots for statistically significant covariates were examined to further explore their influence on the model. Figure 4 below displays the residuals plot.

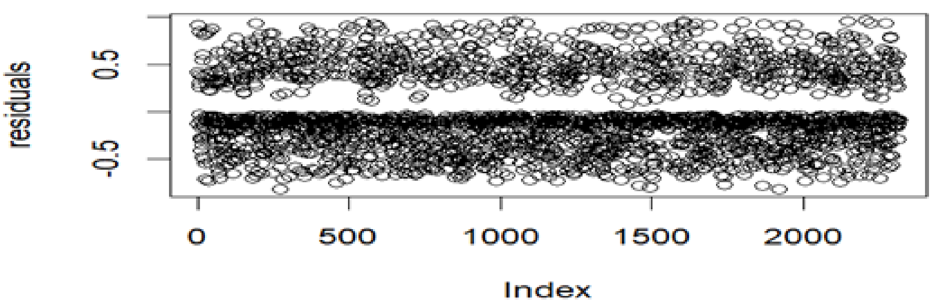

Fig. 4: Residuals Plot for the Fitted Model

The residuals plot displayed in Figure 4 suggests that the model fits the data well. The residuals are symmetrically distributed around zero, show no clear pattern, and exhibit constant variance, indicating no systematic bias in the model's predictions. Ideally, residuals should be centred around zero for the model to be considered well-fitted.

Figure 5 displays the Q-Q plot of the residuals. The plot in Figure 5 indicates that the residuals deviate from normality, showing an S-shaped pattern, which suggests heavier tails than a normal distribution. However, in spatial modelling, particularly with convolution models, residuals are not always expected to follow a normal distribution due to inherent spatial dependencies.

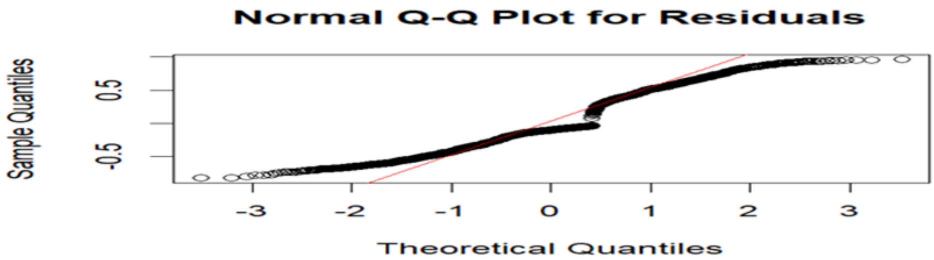

Fig. 5: Normal Q-Q Plot for the Residuals

Several researchers have highlighted that spatial models, such as the convolution model introduce spatial dependency into the residuals, leading to deviations from normality. This is a recognised characteristic in spatial statistics and should be understood as a reflection of the model's ability to capture spatial structure rather than strictly adhering to normality assumptions [34,51,54,55,56].

The global Moran's I statistic for the residuals was -0.0017 (p-value = 0.5373). The non-significant p-value ($p > 0.05$) indicates that the residuals exhibit no statistically significant spatial autocorrelation, suggesting that the convolution model effectively accounted for most of the spatial structure in the data.

Similarly, Geary's C statistic was 0.9925 (p-value = 0.3049), which is close to 1, further confirming the absence of significant spatial autocorrelation in the residuals. The non-significant p-value supports the conclusion that the model adequately captured spatial dependencies. Together, the results of Moran's I and Geary's C tests affirm that the residuals are spatially uncorrelated, reflecting a good model fit.

Additionally, the variogram plot of the residuals (Figure 6) provides further evidence of spatial independence. The lack of spatial correlation in the residuals confirms that the fitted model has effectively addressed the spatial structure within the data.

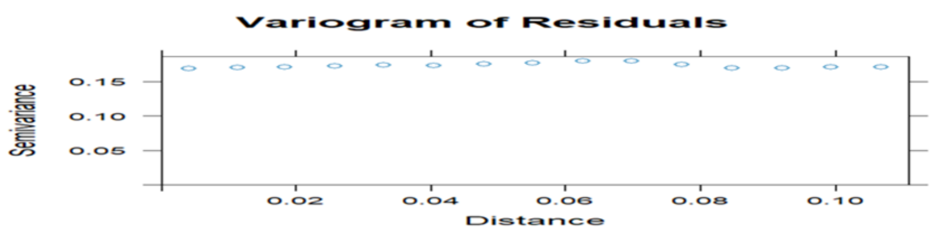

Fig. 6: Variogram Plot for the Residuals

The plot shows a flat or constant semi-variance around 0.15, indicating that the residuals from the fitted convolution model lack spatial autocorrelation, suggesting spatial independence. If the residuals were spatially dependent, such as clustering of positive or negative residuals in space, the semi-variance would increase with distance at some point. However, this pattern was not observed in the plot.

Figure 7 displays the posterior density plots of the statistically significant regression parameters of the structured model. All the plots display smooth curves and single peaks, suggesting stability and proper convergence of the model.

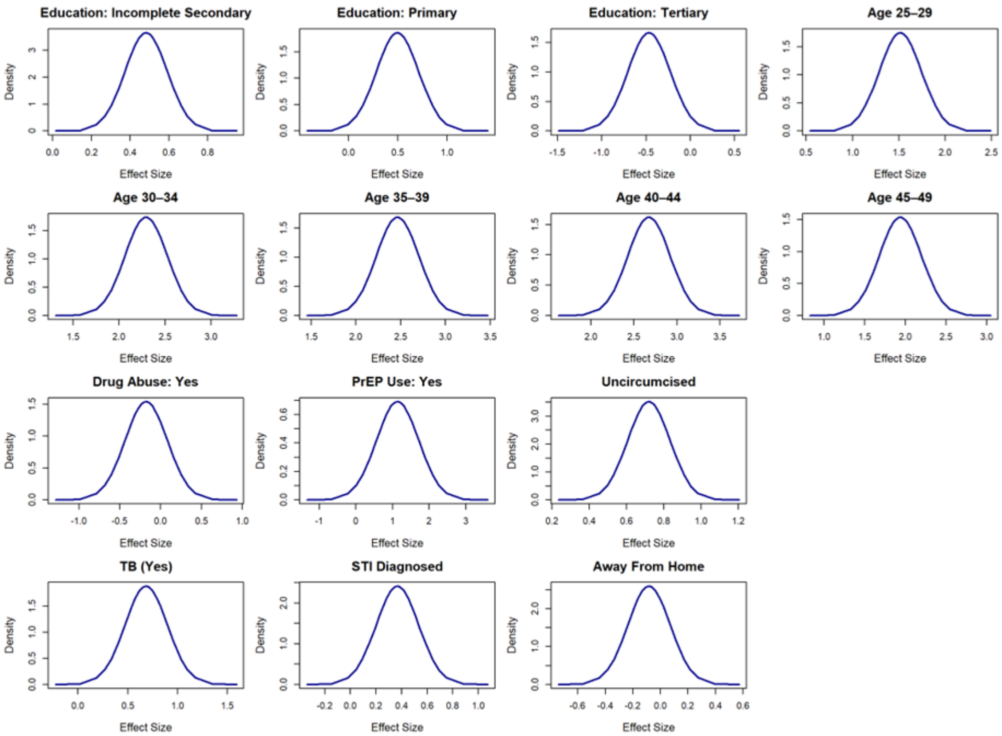

Fig. 7: Posterior Density Plots for the Statistically Significant Coefficients in the Model

Based on the spatial autocorrelation tests applied, the results revealed that the residuals were not spatially autocorrelated. This suggests that the convolution model was appropriate and successfully captured the spatial structure in the data. Further support for the model's adequacy comes from the smooth and unimodal patterns observed in the posterior density plots.

3.2 Discussion

This study explored the spatial disparities and risk factors associated with HIV prevalence among males aged 15–49 in KwaZulu-Natal, South Africa, using a Bayesian convolution model. The findings highlight critical determinants that align with and extend existing knowledge about HIV epidemiology. The overall HIV prevalence of 34% aligns with earlier work [57,58] that positions KZN among regions with the highest HIV prevalence rates globally. Age was a significant predictor of HIV prevalence, with men aged 30–44 being at the highest risk. This is consistent with previous studies that report peak HIV prevalence in similar age groups due to increased sexual activity, multiple partnerships, and cumulative exposure to risk over time [9,17,59,60]. The sharp increase in risk among men aged 30–44 highlights the need for targeted interventions such as scaling up testing and prevention programmes for middle-aged men.

Educational attainment was found to be a key socioeconomic factor, with males having incomplete secondary or primary education at higher risk of HIV infection compared to those with complete secondary or tertiary education. These findings agree with studies linking lower educational levels to reduced access to HIV-related knowledge, prevention resources, and healthcare services [18,19]. Conversely, tertiary education appeared protective, likely due to better awareness and adoption of safer sexual practices. This highlights the importance of integrating HIV education into school curricula, particularly for underserved populations.

Drug abuse was significantly associated with higher odds of HIV infection, a finding consistent with research indicating that substance use is a major driver of risky sexual behaviour, including inconsistent condom use and multiple sexual partners [61,62]. Prevention strategies targeting substance abuse, combined with HIV prevention programmes, are essential to reduce the dual burden of these intertwined public health issues. Unexpectedly, PrEP users had higher odds of HIV positivity, possibly due to adherence challenges, initiation after exposure, or selection bias, where individuals at higher risk are more likely to be prescribed PrEP. These findings align with studies emphasising the importance of adherence counselling and regular monitoring for those on PrEP [63,64].

Regarding male circumcision status, uncircumcised men had significantly higher odds of being HIV-positive compared to circumcised males, consistent with evidence that male circumcision is associated with reduced HIV risk [11,65]. Scaling up voluntary medical male circumcision (VMMC) programmes, particularly in rural and underserved areas, remains a key intervention for HIV prevention. We observed that men diagnosed with TB were at higher risk of HIV infection, reaffirming the strong bidirectional relationship between HIV and TB [66]. This is also consistent with the well-documented association between TB and HIV co-infection, as highlighted in the research by [67,68].

Additionally, our analysis revealed that STIs are significantly associated with increased HIV risk. This is supported by research indicating that individuals diagnosed with STIs have higher odds of HIV infection [11]. This implies that STIs are a marker of risky sexual behaviour and a biological facilitator of HIV transmission. Hence, scaling up STI prevention and treatment services may help mitigate this risk.

Mobility, particularly time spent away from home, was also linked to a higher risk of HIV infection, reinforcing prior research on migration and HIV vulnerability. This aligns with studies suggesting that migration disrupts social networks and increases exposure to high-risk environments [69]. This is also supported by research showing that individuals who travel or are away from home frequently have higher odds of HIV infection [70].

In contrast to some previous studies, we did not find significant associations between HIV infection and main income, marital status, condom use, multiple sexual partners, alcohol use, or length of residence in the community. These discrepancies could be attributed to differences in study design, demographic composition, or regional contexts. Further research is needed to explore these variations.

Figure 2 highlights geographic disparities in HIV prevalence. Higher prevalence areas tend to be concentrated in urban and peri-urban regions, while lower prevalence areas are more widely dispersed. These patterns likely reflect population density, socioeconomic conditions, and healthcare accessibility. Recognising these spatial trends is crucial for designing targeted interventions and optimising resource allocation for HIV prevention and treatment programmes.

High-prevalence areas may benefit from focused strategies, including HIV awareness campaigns, expanded PrEP distribution, and mobile health services. A spatial hotspot near Pietermaritzburg exhibited elevated HIV prevalence, though it did not have statistical significance. However, this pattern suggests the potential need for localised interventions.

Similar spatial analyses in South Africa have identified geographical clusters of high HIV prevalence, often linked to socioeconomic vulnerabilities and disparities in healthcare access [15,71,72,73]. Additionally, studies from other sub-Saharan African settings, including Uganda, have reported significant HIV heterogeneity across different communities, highlighting the role of geographical and population-specific risk factors in shaping the epidemic [74,75]. These findings collectively reinforce the existence of localised HIV heterogeneity and the need for context-specific interventions.

The Bayesian convolution model effectively captured spatial dependencies and heterogeneity, yielding more precise and reliable HIV prevalence estimates across enumeration areas in uMgungundlovu District. By incorporating both structured and unstructured spatial effects, this approach enhances the robustness of epidemiological analyses while accounting for spatial autocorrelation and unmeasured confounders. Our study underscores the advantages of using Bayesian methods in HIV research, offering a more comprehensive understanding of geographical disparities while adjusting for both individual and community-level covariates.

3.3 Conclusion

This study significantly advances spatial epidemiology and HIV research by applying a Bayesian convolution model to analyse HIV prevalence among males aged 15–49 in KwaZulu-Natal, South Africa. Unlike conventional approaches, this method accounts for both spatial heterogeneity and dependence, resulting in more accurate prevalence estimates and a nuanced understanding of geographic disparities. The analysis revealed elevated HIV prevalence in urban and peri-urban regions, with a notable, though not statistically significant, hotspot near Pietermaritzburg. The integration of diverse sociodemographic, behavioural, and biological factors highlighted key risk determinants, including older age (particularly 30–44 years), low educational attainment, lack of circumcision, drug abuse, tuberculosis, sexually transmitted infections (STIs), and mobility.

Conversely, variables such as marital status, alcohol use, condom use, multiple sexual partners, and income source were not significantly associated with HIV status, suggesting that drivers of HIV transmission may differ across regions or contexts. These findings carry important public health implications, underscoring the need for geographically targeted, male-focused HIV interventions. Expanding awareness campaigns, improving access to PrEP, and deploying mobile health services in high-risk areas are critical next steps. Moreover, the strong associations between HIV, TB, and STIs support the adoption of integrated healthcare strategies.

Future research should utilise longitudinal data to better capture temporal patterns of HIV transmission and explore how socioeconomic and structural factors shape spatial disparities. Comparative studies across provinces could offer insight into the generalisability of these patterns. Ultimately, implementing localised, data-driven interventions that address both individual and structural risk factors is essential for reducing HIV prevalence and advancing South Africa's long-term public health goals.

Acknowledgement

We gratefully acknowledge CAPRISA for providing access to the dataset used in this study. We also thank the journal editors and anonymous reviewers for their valuable comments and suggestions that helped improve the quality of this manuscript.

Competing interests

The authors declare that they have no conflicts of interest to disclose.

References

[1] UNAIDS, *Global AIDS Update 2023*, Joint United Nations Programme on HIV/AIDS (UNAIDS), Geneva, 2023.
[2] Human Sciences Research Council (HSRC), *New HIV Survey Highlights Progress and Ongoing Disparities in South Africa's HIV Epidemic*, HSRC Press, Johannesburg, South Africa, 2023.
[3] T. Shand, H. Thomson-de Boor, W. van den Berg, D. Peacock, L. Pascoe, "The HIV blind spot: men and HIV testing, treatment and care in sub-Saharan Africa," *Culture, Health & Sexuality*, **16**(7), 972–985 (2014).
[4] L. Katirayi, T. Maphosa, L. Chilikutali, E. Mhango, A. Bwanali, B. Bvochora, et al., "Understanding gender differences of people with HIV newly diagnosed or returning to care with advanced HIV disease in Malawi: a qualitative study," *BMC Public Health*, **23**, 2382 (2023).
[5] E. Lopez-Varela, O. Augusto, L. Fuente-Soro, R. Sacoor, C. Nhacolo, F. Honwana, et al., "Quantifying the gender gap in the HIV care cascade in southern Mozambique: we are missing the men," *PLoS One*, **16**(2), e0245461 (2021).
[6] M. Cornell, M. Schomaker, D.B. Garone, J. Giddy, C.J. Hoffmann, R. Lessells, et al., "Gender differences in survival among adult patients starting antiretroviral therapy in South Africa: a multicentre cohort study," *PLoS Medicine*, **9**(9), e1001304 (2012).
[7] F. Mosha, V. Muchunguzi, M. Matee, J. Kamugisha, A. Mmbaga, E. Nagu, et al., "Gender differences in HIV disease progression and treatment outcomes among HIV patients one year after starting antiretroviral treatment in Dar es Salaam, Tanzania," *BMC Public Health*, **13**, 38 (2013).
[8] M.H. Shezi, B. Tlou, S. Naidoo, "Knowledge, attitudes, and acceptance of voluntary medical male circumcision among males attending high school in Shiselweni region, Eswatini: a cross-sectional study," *BMC Public Health*, **23**, 349 (2023).
[9] A. Tomita, A. Vandormael, T. Bärnighausen, T. de Oliveira, F. Tanser, "Social determinants of HIV infection clustering in KwaZulu-Natal, South Africa," *AIDS and Behavior*, **21**(8), 2417–2426 (2017).
[10] R. Jewkes, R. Morrell, "Hegemonic masculinity, violence, and gender-based violence: Understanding the intersections," *Journal of Social Issues*, **74**(3), 456–476 (2018).
[11] C.L.J. Ugwu, J.R. Ncayiyana, "Spatial disparities of HIV prevalence in South Africa: Do sociodemographic, behavioral, and biological factors explain this spatial variability?" *Frontiers in Public Health*, **10**, 994277 (2022).
[12] E. Chireshe, R. Chifurira, K. Chinhamu, J.M. Batidzirai, A.B.M. Kharsany, "Spatial analysis of HIV determinants among females aged 15–34 in KwaZulu-Natal, South Africa: a Bayesian spatial logistic regression model," *International Journal of Environmental Research and Public Health*, **22**(3), 446 (2025).
[13] F. Tanser, T. Bärnighausen, A. Dobra, B. Sartorius, "Identifying 'corridors of HIV transmission' in a severely affected rural South African population: a case for a shift toward targeted prevention strategies," *International Journal of Epidemiology*, **47**(2), 537–549 (2018).
[14] L.K. Debusho, N.G. Bedaso, "Bayesian spatial modelling of HIV prevalence in Jimma Zone, Ethiopia," *Diseases*, **11**(1), 46 (2023).
[15] H. Kim, F. Tanser, A. Tomita, A. Vandormael, D.F. Cuadros, "Beyond HIV prevalence: identifying people living with HIV within underserved areas in South Africa," *BMJ Global Health*, **6**, e004089 (2021).
[16] A. Akullian et al., "Large age shifts in HIV-1 incidence patterns in KwaZulu-Natal...", *Proc. Natl. Acad. Sci. U.S.A.*, **118**(28), e2013164118 (2021).
[17] O. Shisana et al., *South African National HIV Prevalence, Incidence and Behaviour Survey, 2012*, HSRC Press, Cape Town (2014).
[18] J.R. Hargreaves et al., "Systematic review exploring time trends in the association between educational attainment and HIV...", *AIDS*, **22**(3), 403–414 (2008).
[19] M. Mabaso et al., "HIV prevalence in South Africa through gender and racial lenses...", *Int. J. Equity Health*, **18**, 167 (2019).

[20] V. Gómez-Rubio, *Bayesian Inference with INLA*, Chapman & Hall/CRC, Boca Raton, FL (2020).
[21] P.A.P. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, **37**(1–2), 17–23 (1950).
[22] L.A. Waller, C.A. Gotway, *Applied Spatial Statistics for Public Health Data*, Wiley, Hoboken, NJ (2004).
[23] W. Hu, K. Mengersen, S. Tong, "Risk factor analysis and spatiotemporal CART model of cryptosporidiosis...", *BMC Infect. Dis.*, **10**, 311 (2010).
[24] R. Haining, G. Li, *Modelling Spatial and Spatial-Temporal Data: A Bayesian Approach*, CRC Press, Boca Raton, FL (2020).
[25] A.D. Cliff, J.K. Ord, *Spatial Processes: Models & Applications*, Pion, London (1981).
[26] L. Anselin, "Local Indicators of Spatial Association—LISA," *Geogr. Anal.*, **27**(2), 93–115 (1995).
[27] R.C. Geary, "The contiguity ratio and statistical mapping," *Incorporated Stat.*, **5**(3), 115–145 (1954).
[28] M. Kulldorff, "Spatial scan statistics: Models, calculations, and applications," in: *Scan Statistics and Applications*, Springer, pp. 303–322 (1999).
[29] S. Banerjee, B.P. Carlin, A.E. Gelfand, *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, Boca Raton, FL (2004).
[30] P. Moraga, *Spatial Statistics for Data Science: Theory and Practice with R*, Chapman & Hall/CRC, Boca Raton, FL (2023).
[31] A. Gelman et al., *Bayesian Data Analysis*, 3rd ed., CRC Press (2013).
[32] J.M. Bernardo, A.F.M. Smith, *Bayesian Theory*, Wiley (2000).
[33] S.J. Press, *Subjective and Objective Bayesian Statistics*, 2nd ed., Wiley (2003).
[34] S. Banerjee, B.P. Carlin, A.E. Gelfand, *Hierarchical Modelling and Analysis for Spatial Data*, 2nd ed., CRC Press (2014).
[35] N. Cressie, C.K. Wikle, *Statistics for Spatio-Temporal Data*, John Wiley & Sons (2011).
[36] A.E. Gelfand, S. Banerjee, *Bayesian Modeling and Analysis of Spatial Data*, CRC Press (2017).
[37] D. Simpson, H. Rue, A. Riebler, T. Martins, S.H. Sørbye, B. Lindgren, et al., "Penalising model component complexity: a principled, practical approach to constructing priors," *Stat. Sci.*, **32**(1), 1–28 (2017).
[38] A. Riebler, S.H. Sørbye, D. Simpson, H. Rue, "An intuitive Bayesian spatial model for disease mapping that accounts for scaling," *Stat. Methods Med. Res.*, **25**(4), 1145–1165 (2016).
[39] C.P. Robert, G. Casella, *Monte Carlo Statistical Methods*, Springer (2004).
[40] W.R. Gilks, S. Richardson, D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC (1995).
[41] S. Brooks, A. Gelman, G.L. Jones, X.-L. Meng, *Handbook of Markov Chain Monte Carlo*, CRC Press (2011).
[42] H. Rue, S. Martino, N. Chopin, "Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (INLA)," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **71**(2), 319–392 (2009).
[43] M. Blangiardo, M. Cameletti, *Spatial and Spatio-Temporal Bayesian Models with R-INLA*, John Wiley & Sons (2015).
[44] F. Lindgren, H. Rue, J. Lindström, "An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **73**(4), 423–498 (2011).
[45] D.M. Blei, A. Kucukelbir, J.D. McAuliffe, "Variational Inference: A Review for Statisticians," *J. Am. Stat. Assoc.*, **112**(518), 859–877 (2017).
[46] M.I. Jordan, Z. Ghahramani, T.S. Jaakkola, L.K. Saul, "An Introduction to Variational Methods for Graphical Models," *Mach. Learn.*, **37**(2), 183–233 (1999).
[47] K.P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press (2012).
[48] D.C. Montgomery, E.A. Peck, G.G. Vining, *Introduction to Linear Regression Analysis*, 5th ed., Wiley (2012).
[49] M.H. Kutner, C.J. Nachtsheim, J. Neter, W. Li, M. Guttman, S. Mantel, et al., *Applied Linear Regression Models*, McGraw-Hill (2004).
[50] N.R. Draper, H. Smith, *Applied Regression Analysis*, 3rd ed., Wiley (2004).
[51] N.A.C. Cressie, *Statistics for Spatial Data (Wiley Series in Probability and Statistics)*, John Wiley & Sons (1993).
[52] C.F. Dormann et al., "Methods to account for spatial autocorrelation in the analysis of species distributional data: A review," *Ecography*, **30**(5), 609–628 (2007).
[53] E.H. Isaaks, R.M. Srivastava, *An Introduction to Applied Geostatistics*, Oxford University Press, Oxford (1989).
[54] R. Haining, *Spatial Data Analysis: Theory and Practice*, Cambridge University Press (2003).
[55] J. Besag, J. York, A. Mollié, "Bayesian image restoration, with two applications in spatial statistics," *Ann. Inst. Stat. Math.*, **43**(1), 1–20 (1991).
[56] D. Gamerman, H.F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed., Chapman and Hall/CRC Press (2006).
[57] G. Ramjee, B. Sartorius, N. Moris, H. Wand, T. Smit, N. Ngcobo, et al., "A decade of sustained geographic spread of HIV infections among women in Durban, South Africa," *BMC Infect. Dis.*, **19**, 500 (2019).
[58] A.B.M. Kharsany, C. Cawood, D. Khanyile, A. Grobler, L.R. McKinnon, N. Samsunder, et al., "Strengthening HIV surveillance in the antiretroviral therapy era: Rationale and design of a longitudinal study to monitor HIV prevalence and incidence in the uMgungundlovu District, KwaZulu-Natal, South Africa," *BMC Public Health*, **15**, 1149 (2015).
[59] A. Pettifor, C. MacPhail, J.P. Hughes, A. Selin, J. Wang, F.X. Gómez-Olivé, et al., "The effect of a conditional cash transfer on HIV incidence in young women in rural South Africa (HPTN 068): A phase 3, randomised controlled trial," *Lancet Glob. Health*, **4**(12) (2018).
[60] K. Govender, S.E. Beckett, G. George, M. Govender, T. Khan, J. Drain, et al., "Factors associated with HIV in younger and older adult men in South Africa: findings from a cross-sectional survey," *BMJ Open*, **9**, e031667 (2019).

[61] J.C. Fisher, H. Bang, S.H. Kapiga, "The association between HIV infection and alcohol use: a systematic review and meta-analysis of African studies," *Sex Transm. Dis.* (2007).

[62] M. Malahleha, F. Laher, A. Dilraj, K. Mngadi, K. Goga, L. Bekker, et al., "Risk factors associated with HIV acquisition in males participating in HIV vaccine efficacy trials in South Africa," *AIDS Behav.*, **27**, 3027–3037 (2023).

[63] R. Eakle, A. Bourne, J. Mbogua, N. Mutanha, H. Rees, "Exploring acceptability of oral PrEP prior to implementation among female sex workers in South Africa," *J. Int. AIDS Soc.*, **20**(Suppl 4):21710 (2017).

[64] L.A. Eaton, D.D. Matthews, L.A. Bukowski, M.R. Friedman, C.J. Chandler, D.L. Whitfield, et al., "Elevated HIV prevalence and correlates of PrEP use among a community sample of Black men who have sex with men," *J. Acquir. Immune Defic. Syndr.*, **79**(3), 339–346 (2018).

[65] B. Auvert, D. Taljaard, E. Lagarde, J. Sobngwi-Tambekou, R. Sitta, A. Puren, "Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: the ANRS 1265 Trial," *PLoS Med.*, **2**(11):e298 (2005).

[66] World Health Organization (WHO), *Global Tuberculosis Report 2019*, Geneva: WHO (2019).

[67] J. Bruchfeld, M. Correia-Neves, G. Källenius, "Tuberculosis and HIV coinfection," *Cold Spring Harb. Perspect. Med.*, **5**(7):a017871 (2015).

[68] I. Navasardyan, R. Miwalian, A. Petrosyan, S. Yeganyan, V. Venketaraman, "HIV-TB coinfection: current therapeutic approaches and drug interactions," *Viruses*, **16**(3):321 (2024).

[69] M.N. Lurie, B.G. Williams, K. Zuma, D. Mkaya-Mwamburi, G.P. Garnett, A.W. Sturm, et al., "The impact of migration on HIV-1 transmission in South Africa: a study of migrant and nonmigrant men and their partners," *Sex Transm. Dis.*, **30**(2), 149–156 (2003).

[70] C.S. Camlin, V. Hosegood, M.-L. Newell, N. McGrath, T. Bärnighausen, R.C. Snow, "Gender, migration, and HIV in rural KwaZulu-Natal, South Africa," *PLoS ONE*, **5**(7):e11539 (2010).

[71] F. Tanser, T. Bärnighausen, G.S. Cooke, M.L. Newell, "Localised spatial clustering of HIV infections in a widely disseminated rural South African epidemic," *Int. J. Epidemiol.*, **38**(4), 1008–1016 (2009).

[72] H. Wand, G. Ramjee, "Targeting the hotspots: Investigating spatial and demographic variations in HIV infection in small communities in South Africa," *J. Int. AIDS Soc.*, **13**, 41 (2010).

[73] H. Wand, C. Whitaker, G. Ramjee, "Geoadditive models to assess spatial variation of HIV infections among women in local communities of Durban, South Africa," *Int. J. Health Geogr.*, **10**, 28 (2011).

[74] L.W. Chang, M.K. Grabowski, R. Ssekubugu, et al., "Heterogeneity of the HIV epidemic in agrarian, trading, and fishing communities in Rakai, Uganda: An observational epidemiological study," *Lancet HIV*, **3**(8) (2016).

[75] A. Kamali, R.N. Nsubuga, E. Ruzagira, U. Bahemuka, G. Asiki, M.A. Price, et al., "Heterogeneity of HIV incidence: A comparative analysis between fishing communities and a neighbouring rural general population, Uganda, and implications for HIV control," *Sex Transm. Infect.*, **92**(6), 447–454 (2016).

Exaverio Chireshe is a second-year PhD student in Statistics at the University of KwaZulu-Natal. His research interests include multivariate analysis, spatial analysis, biostatistics, and time series. He focuses on the application of Bayesian and hierarchical models to public health and epidemiological data, with a particular interest in supporting data-driven policy and decision-making in resource-limited settings.

Retius Chifurira is an Associate Professor of Statistics at the University of KwaZulu-Natal, with a PhD from the University of the Free State and over 20 years of academic experience. His expertise lies in multivariate analysis, biostatistics, financial time series analysis, and econometric modelling. He has published widely on topics such as joint survival models and value-at-risk estimation. Prof. Chifurira is actively involved in promoting statistics education in Africa, having chaired a session at the 64th ISI World Statistics Congress and serving as an interim executive member of the Zimbabwe Statistics Association. He is also a member of the South African Statistical Association and has supervised numerous postgraduate students.

Jesca M Batidzirai is an experienced biostatistician with a strong background in health research and has extensive experience lecturing at undergraduate and postgraduate levels. She has played a pivotal role in shaping the next generation of researchers and professionals in the field. To date, she has successfully supervised 4 MSc students across a range of biostatistics topics and continues to actively mentor emerging scholars and junior researchers. She is currently supervising 2 PhD and 4 MSc students. Her expertise lies in Time-to-event analysis and Spatial Analysis, with much of her work embedded in interdisciplinary and collaborative research. Jesca is passionate about leveraging statistical science to address critical health challenges and is deeply committed to fostering academic growth and promoting evidence-based research through rigorous statistical methodologies. Her research is primarily dedicated to maternal, child, and mental health, with a strong emphasis on enhancing well-being and promoting positive health outcomes.

Knowledge Chinhamu is an Associate Professor at the University of KwaZulu-Natal, South Africa. He received his PhD in Statistics from the University of the Free State, South Africa. His research interests include applied statistics, time series analysis, financial modelling, and biostatistics. He has published research articles in reputable international journals and is a referee for statistical journals.

Ayesha BM Kharsany is a Professor in Medical Microbiology at the University of KwaZulu-Natal, South Africa. She completed her PhD and obtained her Masters in Epidemiology from Columbia University, New York, USA. She is a member of The Academy of Science of South Africa (ASSA). As a Senior Scientist and an epidemiologist at CAPRISA, she has extensive research, professional and leadership experience and plays a leading role in the design, regulatory requirements, operations, and analysis of HIV and STI epidemiological studies. Her research and expertise focus are on understanding the evolving HIV epidemic in South Africa, factors influencing HIV acquisition in young women, monitoring HIV prevalence and incidence in association with the scale-up of prevention efforts in "real world", non-trial settings and understanding of the cycle of HIV transmission in generalized hyperendemic HIV epidemic settings that are heterosexually driven. She is committed to training and mentoring the next generation of scientists including scientific leadership to advance the epidemiological, behavioural and social science research in the field of HIV and STIs.