Spatial Patterns of HIV Transmission in Kenya: A Bayesian Analysis

Mulati Omukoba Nyukuri*[1], Bonface Kwach[1]

[1]Department of Mathematics, Kibabii University, P.O. Box 1699–50200, Bungoma, Kenya

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.101000545

Received: 21 October 2026; Accepted: 26 November 2026; Published: 07 November 2026

ABSTRACT

Kenya's HIV epidemic is highly uneven across its 47 counties, yet programme planning still relies largely on national or regional averages. This study estimated county-level HIV transmission potential, identified transmission hotspots and described inter-county connectivity and seasonal variation in transmission. A quantitative ecological modelling design was used, drawing on spatial epidemic theory and hierarchical Bayesian theory. County-month surveillance data for 15 purposively selected counties in five regions (2015–2024) were compiled from the National AIDS Control Council, the Kenya Health Information System, national population-based HIV surveys and the Kenya National Bureau of Statistics, and analysed with an adaptive spatial hierarchical Bayesian SEIR model fitted by Hamiltonian Monte Carlo in Stan; county-specific (local) reproduction numbers and a system reproduction number were derived from the next-generation matrix. Reproduction numbers ranged from 0.3 in Wajir to 3.2 in Homa Bay. Homa Bay (3.2), Siaya (3.1) and Kisumu (2.8) formed the only high-transmission cluster; ten counties were of medium and two north-eastern counties of low transmission. The posterior spatial correlation range was 85.3 km (95% CrI 62.1–108.5). Connectivity within the western cluster was strong and stable (mean weight 0.86, CV 1.4%), whereas the Nairobi–Mombasa corridor was weaker and the most variable (mean 0.39, CV 9.7%), reaching its lowest value in 2020. Transmission peaked in October in high-burden counties and in July in low-burden north-eastern counties. The study contributes by moving Kenyan HIV spatial analysis from mapping prevalence to estimating transmission potential and time-varying connectivity within a single framework. The results indicate that transmission potential is concentrated in one tightly connected Lake Victoria cluster, implying that the three counties should be planned jointly, that prevention resources should be allocated by transmission potential, and that the exploratory seasonal pattern should be monitored before outreach is timed to it. Because the design is ecological, associational and limited to 15 counties, the findings describe county-level patterns within the study area and should not be generalised to other counties or to individuals without further evidence.

Indexed Terms – HIV, Spatial Epidemiology, Bayesian Hierarchical Model, SEIR Model, Reproduction Number, Hotspots, Kenya

INTRODUCTION

HIV remains one of the leading public health challenges in sub-Saharan Africa, and Kenya has one of the largest epidemics in the world. About 1.3 million adults were living with HIV at the time of the 2018 Kenya Population-based HIV Impact Assessment (KENPHIA), which surveyed adults aged 15–64 years, with a national adult prevalence of 4.9% (National AIDS and STI Control Programme [NASCOP], 2022). This national figure conceals very large differences between counties. The same survey recorded adult prevalence of 19.6% in Homa Bay and 17.5% in Kisumu, against less than 0.2% in parts of the former North Eastern region (NASCOP, 2022). Few countries in the region show a steeper sub-national gradient (Dwyer-Lindgren et al., 2019; Cuadros et al., 2017).

This geographic concentration has direct consequences for policy. Prevention and treatment resources are limited, and a uniform allocation spends heavily where transmission is already low. Modelling work in Kenya has shown that prioritising the places and populations with the greatest need can avert substantially more infections for the same budget (Anderson et al., 2014). The Kenya AIDS Strategic Framework II accordingly

commits the national response to a location-population approach that directs resources to where new infections occur (National AIDS Control Council [NACC], 2020). Under Kenya's devolved system, in which each county plans and budgets its own health services, putting that commitment into practice requires county-level evidence on where transmission is concentrated and how counties are linked.

Prevalence alone is a poor guide to where new infections will occur. It reflects past incidence, survival on treatment and migration as well as current transmission (Tanser et al., 2009). Earlier spatial studies in Kenya mapped prevalence and its covariates using Bayesian methods (Ngesa et al., 2014; Otiende et al., 2019), but they did not link those patterns to a transmission mechanism that estimates how many secondary infections each case generates in each county, or how counties are connected by population movement.

These considerations raise the central empirical question addressed by this study: how is HIV transmission potential distributed across Kenyan counties, and how are high-transmission counties connected to each other and to the rest of the country? The study forms part of a broader investigation that developed an adaptive spatial hierarchical Bayesian SEIR model for HIV in Kenya (Mulati et al., 2025a), established its mathematical properties (Mulati et al., 2025b) and validated its predictive performance (Mulati, 2026). The present paper reports the substantive epidemiological findings. Its specific objectives were:

1. to estimate county-level HIV transmission potential, measured by the county reproduction number;
2. to identify HIV transmission hotspots and describe the connectivity between counties over time;
3. to describe seasonal variation in HIV transmission across counties of different burden.

LITERATURE REVIEW

Conceptualising Spatial Patterns of HIV Transmission

Spatial patterns of HIV transmission refer to the way in which new infections, and the potential for further infections, are distributed across geographic units and change over time. Three dimensions are relevant for programme planning. The first is the level of transmission potential in each area, commonly summarised by the reproduction number, the expected number of secondary infections generated by one infected person (Diekmann et al., 1990; van den Driessche & Watmough, 2002). The second is clustering: whether areas of high transmission lie next to one another and form hotspots. The third is connectivity: the extent to which infection in one area depends on infection in others through population movement and sexual partnerships that span places.

Spatial Heterogeneity of HIV in Sub-Saharan Africa

HIV in sub-Saharan Africa clusters at every spatial scale that has been studied. In rural KwaZulu-Natal, Tanser et al. (2009) found sharply bounded clusters of infection that individual risk factors did not explain. Continental mapping has since shown that variation within countries often exceeds variation between them, with eastern Africa's highest-burden districts concentrated along the Lake Victoria basin (Dwyer-Lindgren et al., 2019). Cuadros et al. (2017) argued that maps of localised variation provide effective information for targeted prevention and control. These findings support targeting, but most of the underlying analyses describe prevalence rather than the potential for further transmission.

Bayesian Spatial Modelling of HIV in Kenya

Hierarchical Bayesian models are the standard approach to mapping disease across small areas. They borrow strength between neighbouring areas and quantify uncertainty at each level of the model (Besag et al., 1991; Banerjee et al., 2014; Lawson, 2018). In Kenya, Ngesa et al. (2014) used Bayesian semi-parametric spatial models to describe variation in HIV prevalence from national survey data, and Otiende et al. (2019) modelled the joint spatio-temporal risk of HIV and tuberculosis by county. Both confirmed a strong spatial gradient with the highest risk in the Lake Victoria counties. However, these models are statistical rather than mechanistic: they describe where risk is high but do not estimate transmission parameters or reproduction numbers.

Population Mobility and Connectivity

Population movement links places epidemiologically. Using mobile phone data, Wesolowski et al. (2012) showed that travel between Kenyan regions shaped the spread of malaria, establishing methods later adapted to other infections. Mobility is strongly associated with HIV risk across eastern Africa and differs by gender (Camlin et al., 2019). Despite this evidence, county-level HIV models in Kenya have rarely represented connectivity between counties explicitly, and where they have, connectivity has typically been treated as fixed over time.

Policy and Institutional Context of the HIV Response in Kenya

Kenya's HIV response operates within the devolved system created by the Constitution of Kenya, 2010, which established 47 county governments and, under the Fourth Schedule, assigned county health services to the counties while leaving health policy and national referral facilities with the national government (Republic of Kenya, 2010). Counties finance health services mainly from their equitable share of nationally raised revenue, which is divided among counties according to a formula recommended by the Commission on Revenue Allocation and approved by Parliament, supplemented by own-source revenue and conditional grants. The County Governments Act, 2012 requires each county to prepare a County Integrated Development Plan, and the Public Finance Management Act, 2012 governs county budgeting and the management of grants from development partners, in which the National Treasury plays a central role (Republic of Kenya, 2012a, 2012b). The Intergovernmental Relations Act, 2012 established the Council of Governors as the main forum for coordination among counties and between counties and the national government (Republic of Kenya, 2012c).

Within this structure, the National AIDS Control Council coordinates the multisectoral response and sets national strategy through the Kenya AIDS Strategic Framework, while the National AIDS and STI Control Programme in the Ministry of Health sets clinical and programme guidelines; county health departments, through county AIDS and STI coordinators, translate national strategy into County AIDS Strategic Plans and deliver services, reporting routine data to the Kenya Health Information System, a national platform built on DHIS2 (NACC, 2020; Karuri et al., 2014). The HIV and AIDS Prevention and Control Act, 2006 provides the legal basis for HIV prevention, testing and the protection of people living with HIV (Republic of Kenya, 2006). A large share of HIV programme and commodity financing has historically come from external partners, chiefly the United States President's Emergency Plan for AIDS Relief and the Global Fund, and the Strategic Framework commits the country to increasing domestic financing (NACC, 2020). Two features of this arrangement matter for the present study. First, because each county plans and budgets separately, an epidemic that crosses county boundaries has no single planning authority. Second, as national and county budgets take on a larger share of the response, the case for allocating limited resources according to where transmission is concentrated becomes stronger. Regional economic blocs such as the Lake Region Economic Bloc, which includes Homa Bay, Kisumu and Siaya, offer an existing vehicle for joint planning across county lines.

Theoretical Anchorage

This study draws on two complementary theories. Spatial epidemic theory, rooted in the compartmental framework of Kermack and McKendrick (1927) and its extension to HIV (Anderson & May, 1991), holds that the force of infection in an area depends on the number of infectious people both within it and in the areas to which it is connected. It implies that transmission potential can be quantified for each area through the next-generation matrix, and that areas with a reproduction number above one sustain transmission while those below one do not (Diekmann et al., 1990; van den Driessche & Watmough, 2002). Hierarchical Bayesian theory provides the statistical basis for estimating such a model from imperfect data, by treating county parameters as draws from shared distributions and representing unexplained spatial dependence through a structured random process (Banerjee et al., 2014). Together, the two theories justify estimating county transmission potential with full uncertainty while allowing information to be shared across connected counties. Spatial epidemic theory has two further implications that are central to this study. First, in a connected system, an area with a reproduction number below one may still record new infections because infection is repeatedly imported from connected areas, so low local transmission potential is not the same as low risk. Second, interventions in a high-transmission area can reduce infections in the areas connected to it, so the benefits of action are not confined within administrative boundaries.

Both theories have limitations that shape how the results should be read. Spatial epidemic theory treats each area as internally well mixed, whereas HIV transmission is structured by sexual networks, key populations and sub-county hotspots (Keeling & Rohani, 2008; Tanser et al., 2009). The threshold interpretation of the reproduction number was developed for infections spreading in an untreated population; in a mature, treated epidemic such as Kenya's, a value above one indicates that transmission is sustained under current treatment coverage rather than that a new epidemic will grow. Connectivity between areas is also usually represented through distance-decay or gravity assumptions that are only partly observed. Hierarchical Bayesian theory, for its part, gains precision by shrinking county estimates towards shared means, which may understate the most extreme counties when data are sparse. The present study responds to these limitations by allowing connectivity to vary over time, by interpreting $\mathcal{R}$ as a control reproduction number and by reporting full posterior uncertainty, but it does not remove them.

Research Gap

While prior studies have mapped HIV prevalence across Kenya (Ngesa et al., 2014; Otiende et al., 2019) and shown the value of geographic targeting (Anderson et al., 2014), few have estimated county-level transmission potential within a framework that also represents time-varying connectivity between counties and seasonal variation in transmission. This study addresses that gap.

Conceptual Framework

Figure 1 presents the conceptual framework that guided the study. The independent variables are spatial connectivity between counties, temporal and seasonal dynamics of transmission, and county epidemiological and programme characteristics. They are linked to the dependent variable, spatial patterns of HIV transmission, through the adaptive spatial hierarchical Bayesian SEIR model, which serves as the intervening variable; the framework describes how the study variables are related rather than asserting causal effects. Data quality and coverage are treated as moderating variables, because they condition how accurately the underlying patterns can be estimated.

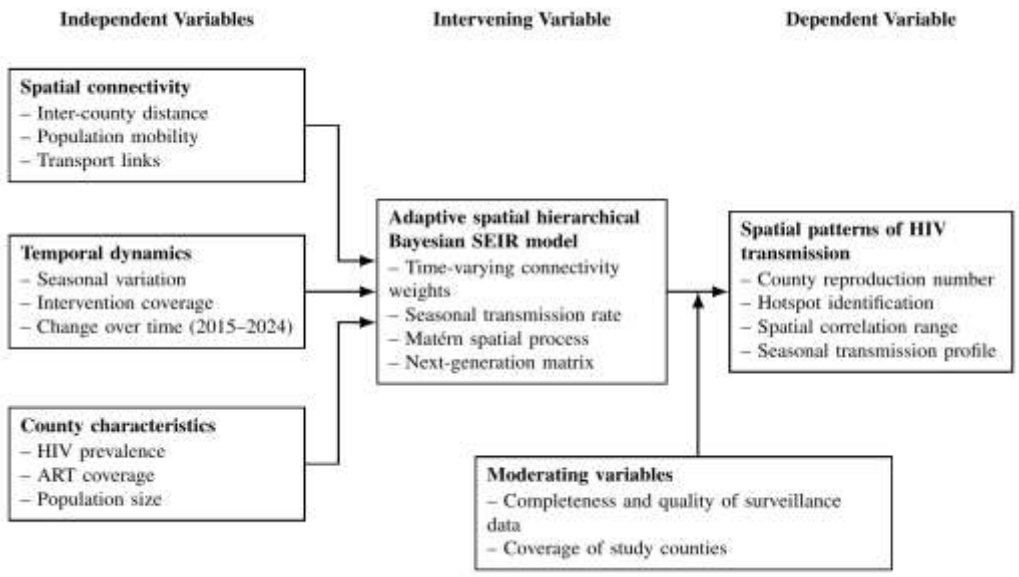

Figure 1: Conceptual Framework

Source: Researcher (2026)

RESEARCH METHODOLOGY

Research Design

The study adopted a quantitative, ecological modelling design within a positivist research philosophy. The unit of analysis was the county-month. A spatial transmission model was fitted to longitudinal county data, and its

estimated parameters were used to describe transmission potential, clustering, connectivity and seasonality. Because the data are aggregated at county level, the analysis describes patterns across counties and cannot be used to draw inferences about individuals.

Study Area and Selection of Counties

The study covered 15 of Kenya's 47 counties, selected purposively to span the national range of HIV prevalence, population density, health-care access and mobility. The selected counties fall in five regions, as shown in Table 1. Because selection was purposive rather than random, the 15 counties are not a statistically representative sample of Kenya's counties; the results apply to the study counties and can be extended to other counties only with caution.

Table 1: Study Counties by Region

Region	Counties	Epidemiological context
Western	Homa Bay, Siaya, Kisumu	High-burden Lake Victoria cluster
Coast	Mombasa, Kilifi, Kwale	Urban and peri-urban coastal corridor
Central	Nairobi, Kiambu, Nakuru	Urban and peri-urban
Eastern	Machakos, Kitui, Meru	Predominantly rural
Northern and North Eastern	Turkana, Garissa, Wajir	Arid and semi-arid, international borders

Data Sources

Secondary data for 2015–2024 were drawn from four sources (Table 2). The analytic dataset comprised 1,800 county-months (15 counties × 120 months). Data were extracted with a standardised data extraction template applied identically to every county. The template recorded the variables listed in Table 2: new HIV infections, people living with HIV and antiretroviral therapy coverage (Kenya HIV Estimates); HIV testing, treatment outcomes and AIDS-related deaths (Kenya Health Information System); prevalence and testing coverage for each survey round (KAIS, KENPHIA); and population by age and sex together with mobility indicators (Kenya National Bureau of Statistics). Facility records were summed to county-months. Before modelling, each series was checked for completeness, for internal consistency (for example, numbers on treatment not exceeding people living with HIV) and consistency between sources and for reporting artefacts such as abrupt changes linked to changes in reporting systems. Missing county-months were not imputed beforehand; they were treated as unknown quantities and estimated within the Bayesian model, so that their uncertainty is carried through to all results. Reported counts were modelled on their original scale with a negative binomial likelihood, with county population entering as an exposure term (see Model Specification). In sensitivity analysis, the model's predictive performance remained acceptable with up to 20% of observations missing (Table 4). In total, 6% of county-months had at least one missing value (14% for the northern counties).

The variables differ in definition and temporal resolution. New HIV infections, people living with HIV and antiretroviral therapy coverage (the percentage of people living with HIV on treatment) are annual county estimates produced by the National AIDS Control Council with the Spectrum model. HIV testing (the number of people tested), treatment outcomes (the numbers newly started on and currently receiving antiretroviral therapy) and AIDS-related deaths are monthly facility counts reported through the Kenya Health Information System, which were summed across facilities to give county-month totals. Prevalence (the percentage of adults testing HIV-positive) and testing coverage are point estimates for each survey round. Population by age and sex comes from the 2019 census and Kenya National Bureau of Statistics projections. The likelihood was defined on county HIV case counts, linked to the modelled number of people living with untreated HIV or AIDS (see Model Specification). Annual NACC estimates were assigned to their reference year and survey prevalence to its survey round; neither was interpolated to monthly values.

Table 2: Data Sources and Variables

Source	Variables	Period	Resolution
National AIDS Control Council (Kenya HIV Estimates)	New infections, people living with HIV, ART coverage	2015–2024	County
Kenya Health Information System	HIV testing, treatment outcomes, AIDS-related deaths	2015–2024	Facility, aggregated to county
National population-based HIV surveys (KAIS, KENPHIA)	Prevalence, testing coverage	Survey rounds	County
Kenya National Bureau of Statistics	Population by age and sex; mobility indicators	2015–2024	County

Research authorisation was obtained from the National Commission for Science, Technology and Innovation (NACOSTI; permit NACOSTI/P/25/4173521), and the proposal was approved by the Kibabii University School of Graduate Studies. The study used only aggregated, de-identified data and complied with Kenya's Data Protection Act, 2019.

Model Specification

The population of each county i was divided into susceptible (S_i), exposed (E_i), HIV-infected (I_i), AIDS (A_i) and on-treatment (R_i) compartments (Mulati et al., 2025a).

The transmission dynamics in county i are described by the system

$$\frac{dS_i}{dt} = \Lambda_i - \lambda_i(t)S_i - \mu S_i, \quad (1a)$$

$$\frac{dE_i}{dt} = \lambda_i(t)S_i - (\sigma + \mu)E_i, \quad (1b)$$

$$\frac{dI_i}{dt} = \sigma E_i - (\gamma + \rho + \mu)I_i, \quad (1c)$$

$$\frac{dA_i}{dt} = \rho I_i - (\alpha + \tau + \mu)A_i, \quad (1d)$$

$$\frac{dR_i}{dt} = \gamma I_i + \tau A_i - \mu R_i, \quad (1e)$$

where Λ_i is recruitment into the sexually active population, σ the rate of progression from exposure to infectiousness, γ the treatment initiation rate, ρ the rate of progression to AIDS, τ the treatment rate among AIDS cases, α AIDS-related mortality and μ background mortality.

The compartmental structure follows the natural history of HIV under treatment. The exposed compartment E represents the period between acquisition and established infection that is detectable in routine data; it is treated as not contributing to onward transmission at population level. This is a simplification, because acute HIV infection is highly infectious (Hollingsworth et al., 2008), but the acute phase is short relative to the course of infection and the progression rate σ is correspondingly fast. The compartment I represents chronic, untreated infection, the main source of transmission in the model. The compartment A represents advanced disease, whose relative infectiousness η allows for the higher viral load of late-stage infection (Hollingsworth et al., 2008) together with reduced sexual activity, and which carries additional AIDS-related mortality α. The compartment R contains people on antiretroviral therapy; they are assumed not to transmit, consistent with evidence that sustained viral suppression prevents sexual transmission (Cohen et al., 2016; Rodger et al., 2019). Treatment is initiated from both I (rate γ) and A (rate τ), reflecting entry into care at earlier and later stages of disease. People on treatment are not returned to the infectious compartments, so treatment interruption and failure are not represented. Because γ enters the transition matrix V, estimated transmission potential depends directly on treatment coverage.

The force of infection in county i, which allows for infection arising from connected counties, is

$$\lambda_i(t) = \beta_i(t) \sum_{j=1}^{n} w_{ij}\,(t) \frac{I_j(t) + \eta A_j(t)}{N_j(t)}, \tag{2}$$

where $w_{ij}(t)$ is the connectivity weight from county j to county i at time t, η is the relative infectiousness of people with AIDS and N_j is the population of county j.

The connectivity weights and the transmission rate vary over time according to

$$w_{ij}(t) = w_{ij}^0 \exp\left(-\delta(t)\, d_{ij}^{\kappa}\right) \psi\left(X_{ij}(t)\right), \tag{3a}$$

$$\beta_i(t) = \beta_{i0}\left[1 + \beta_{i1}\cos\left(\frac{2\pi t}{12}\right) + \beta_{i2}\sin\left(\frac{2\pi t}{12}\right)\right], \tag{3b}$$

where w_{ij}^0 is the baseline connectivity, d_{ij} the distance between county centroids, $\delta(t)$ a time-varying decay rate, κ a distance-decay exponent that controls how quickly connectivity declines with distance, $\psi(\cdot)$ a function of mobility and infrastructure covariates, and β_{i1}, β_{i2} the seasonal coefficients.

Connectivity weights are partly informed by data and partly estimated by the model. The baseline weights w_{ij}^0 were calculated from the physical proximity of counties combined with county-level population movement indicators from the Kenya National Bureau of Statistics (Mulati, 2025). The time-varying decay rate followed $\delta(t) = \delta_0 + \sum_k \delta_k X_k(t) + \zeta\delta(t-1)$, where $X_k(t)$ are time-varying mobility and health-care access covariates from the Kenya National Bureau of Statistics and the Kenya Health Information System, ζ is an autoregressive coefficient, and all coefficients were estimated within the model (Mulati, 2025). (The decay rate is written δ to distinguish it from the AIDS-related mortality rate α.) No direct origin-destination data, such as mobile-phone records, were available. The year-to-year variation in connectivity reported below is therefore a model-based estimate driven by these covariates and by the fit to surveillance data, not an observed flow.

The effect of alternative connectivity and structural assumptions was assessed by refitting the model under alternative specifications (see Data Analysis); the results are reported in Table 4.

The model was embedded in a three-level Bayesian hierarchy (Mulati, 2025). At the observation level, county HIV case counts $Y_i(t)$ followed a negative binomial distribution, $Y_i(t) \sim \text{NegBin}(m_i(t), \phi)$, with $\log m_i(t) = \log N_i(t) + \log\{[I_i(t) + A_i(t)]/N_i(t)\}$. The expected count is therefore county population multiplied by the modelled proportion of the population living with untreated HIV or AIDS, so population enters as an exposure term, the epidemic model enters through that proportion, and ϕ absorbs overdispersion. At the process level, county effects followed a Gaussian process with Matérn covariance with range ℓ, smoothness ν and variance σ_θ^2. At the parameter level, hyperparameters were assigned weakly informative priors.

The priors were $\beta_{i0} \sim \text{Gamma}(a_\beta, b_\beta)$, $\sigma_\theta^2 \sim \text{Inverse-Gamma}(a_\sigma, b_\sigma)$, $\ell \sim \text{Inverse-Gamma}(a_\ell, b_\ell)$ and $\nu \sim \text{Uniform}(0.5,2.5)$, with $a_\beta = 2$, $b_\beta = 16.4$, $a_\sigma = 2.5$, $b_\sigma = 0.68$, $a_\ell = 2.5$ and $b_\ell = 128$. These give prior means of approximately 0.122 for β_{i0}, 0.453 for σ_θ^2 and 85.3 km for ℓ. The priors were diffuse: their 95% intervals were 0.015–0.340 for the baseline transmission rate, 0.11–1.64 for the spatial variance and 20–308 km for the spatial range, about five to six times wider than the corresponding posterior credible intervals (Table 3). The data therefore determined the precision of these estimates.

The numerical coincidence between these prior means and the posterior means in Table 3 is acknowledged; the prior specification is set out in Appendix A. The alternative-prior sensitivity analysis in Table 4 shows that predictive error changed by about 5%, indicating that the main results are not driven by the choice of priors.

Parameter identifiability was examined through the eigenvalues of the Fisher information matrix (Mulati, 2025). Only the leading few parameter directions had eigenvalues above the identifiability threshold; the remainder were weakly constrained by the data, indicating a degree of over-parameterisation relative to 15 counties. The baseline transmission rate, the spatial variance and the spatial range, which also dominated the variance-based sensitivity analysis (Table 4), were estimated with narrow credible intervals (Table 3). Parameters that are numerous relative to the data, such as individual pairwise connectivity weights and seasonal coefficients, are

likely to be informed partly by the hierarchical structure and priors, and results that depend on them individually are interpreted with caution.

The reproduction numbers were derived with the next-generation method (Diekmann et al., 1990; van den Driessche & Watmough, 2002). With the infected compartments ordered as $x = (E_1, I_1, A_1, \dots, E_n, I_n, A_n)^\top$, the infected subsystem of Equation (1) is written as

$$\frac{d\mathbf{x}}{dt} = \mathcal{F}(\mathbf{x}) - \mathcal{V}(\mathbf{x}), \tag{4a}$$

where $F(x)$ contains the rates at which new infections appear and $V(x)$ contains all other transfers into and out of the infected compartments. New infections enter only the exposed compartments, so that, writing $k_1 = \sigma + \mu$, $k_2 = \gamma + \rho + \mu$ and $k_3 = \alpha + \tau + \mu$,

$$\mathcal{F}_{E_i} = \beta_i \sum_{j=1}^{n} w_{ij} S_i \frac{I_j + \eta A_j}{N_j}, \quad \mathcal{F}_{I_i} = \mathcal{F}_{A_i} = 0, \tag{4b}$$

$$\mathcal{V}_{E_i} = k_1 E_i, \quad \mathcal{V}_{I_i} = -\sigma E_i + k_2 I_i, \quad \mathcal{V}_{A_i} = -\rho I_i + k_3 A_i. \tag{4c}$$

The Jacobians $F = \partial F / \partial x$ and $V = \partial V / \partial x$ were evaluated at the disease-free equilibrium, at which $E_i = I_i = A_i = R_i = 0$ and $S_i^0 = \Lambda_i / \mu$. The only non-zero entries of F lie in the rows of the exposed compartments:

$$\frac{\partial \mathcal{F}_{E_i}}{\partial I_j} = \beta_i w_{ij} \frac{S_i^0}{N_j^0}, \quad \frac{\partial \mathcal{F}_{E_i}}{\partial A_j} = \eta \beta_i w_{ij} \frac{S_i^0}{N_j^0}. \tag{4d}$$

Because counties are linked only through infection and not through progression or treatment, V is block-diagonal, $V = \mathrm{diag}(V_0, \dots, V_0)$, with one 3×3 block for each county:

$$V_0 = \begin{pmatrix} k_1 & 0 & 0 \\ -\sigma & k_2 & 0 \\ 0 & -\rho & k_3 \end{pmatrix}. \tag{4e}$$

V_0 is lower-triangular with positive diagonal and non-positive off-diagonal entries, so it is a non-singular M-matrix and its inverse exists and is non-negative. The inverse follows by forward substitution:

$$V_0^{-1} = \begin{pmatrix} \frac{1}{k_1} & 0 & 0 \\ \frac{\sigma}{k_1 k_2} & \frac{1}{k_2} & 0 \\ \frac{\sigma\rho}{k_1 k_2 k_3} & \frac{\rho}{k_2 k_3} & \frac{1}{k_3} \end{pmatrix}. \tag{4f}$$

Each entry of V_0^{-1} has a direct interpretation. The (I, E) entry is the probability of progressing from the exposed to the infectious stage, σ/k_1, multiplied by the mean time spent in I, $1/k_2$; the (A, E) entry is the probability of passing through I to A, multiplied by the mean time spent in A, $1/k_3$. Multiplying the E_i row of F by the E_j column of V^{-1} gives the expected number of new infections in county i generated by one person newly infected in county j, and hence the next-generation matrix and reproduction numbers

$$K_{ij} = \beta_{i0} w_{ij}^0 \frac{S_i^0}{N_j^0} \cdot \frac{\sigma}{\sigma + \mu} \cdot \frac{1}{\gamma + \rho + \mu} \left(1 + \frac{\eta\rho}{\alpha + \tau + \mu}\right), \tag{4g}$$

$$\mathcal{R}_i = K_{ii}, \quad \mathcal{R}_{\mathrm{sys}} = r(K) \geq \max_i \mathcal{R}_i. \tag{4h}$$

Because new infections enter only the exposed compartments, the next-generation matrix FV^{-1} for the $3n$ infected compartments (E_i, I_i, A_i) reduces to an $n \times n$ matrix K, whose element K_{ij} is the expected number of new infections in county i generated by one infected person in county j over the whole infectious period (Diekmann et al., 2010; van den Driessche & Watmough, 2002). In Equation (4g), $\sigma/(\sigma + \mu)$ is the probability of progressing from the exposed to the infectious stage, $1/(\gamma + \rho + \mu)$ is the mean duration of untreated chronic infection, and $\eta\rho/(\alpha + \tau + \mu)$ adds the contribution of the AIDS stage, weighted by its relative infectiousness η. For a single

isolated county ($n = 1$), $S_i^0/N_i^0 = 1$ at the disease-free equilibrium and the expression reduces to $\beta_0/[(\sigma+\mu)(\gamma+\rho+\mu)] \cdot [1 + \eta\rho/(\alpha+\tau+\mu)]$. K was evaluated at the baseline connectivity weights w_{ij}^0, at the annual mean transmission rate β_{i0} (the seasonal terms in Equation (3) average to zero over a year) and at the reference susceptible population S_i^0, so the reproduction numbers summarise transmission potential over a typical year rather than in a particular month.

Two quantities are distinguished. The county-specific (local) reproduction number, $\mathcal{R}_i = K_{ii}$, is the number of secondary infections that one infected person generates within their own county through within-county connectivity w_{ii}^0; it is the quantity reported for each county in Table 5 and used to classify counties. The system reproduction number, $\mathcal{R}_{\text{sys}} = r(K)$, where $r(\cdot)$ denotes the spectral radius, is the reproduction number of the connected 15-county system as a whole and determines whether the disease-free equilibrium of the full system is stable (Mulati et al., 2025b). Because K is non-negative, its spectral radius is at least as large as its largest diagonal element, so $\mathcal{R}_{\text{sys}} \geq \max_i \mathcal{R}_i$. Local values therefore show where transmission potential is concentrated, while the system value governs the connected epidemic as a whole. Because the treatment rates γ and τ enter in the denominators, both are control reproduction numbers, measuring transmission potential under prevailing treatment coverage. Both were computed for every posterior draw, giving a posterior distribution and 95% credible interval for each county.

Data Analysis

The model was fitted by Hamiltonian Monte Carlo using the No-U-Turn Sampler in Stan (Carpenter et al., 2017; Hoffman & Gelman, 2014), with data preparation in R version 4.2.0. Convergence was assessed using the Gelman–Rubin statistic $(\hat{R} < 1.05)$ and effective sample sizes (Gelman & Rubin, 1992). Counties were classified by the posterior mean of the local reproduction number $\mathcal{R}$ as high ($\mathcal{R} > 2$), medium ($1 \leq \mathcal{R} \leq 2$) or low ($\mathcal{R} < 1$) transmission. A transmission hotspot was defined as a high-transmission county within the posterior correlation range ℓ of at least one other high-transmission county, with a connectivity weight above 0.5 between them. Temporal variability in connectivity was summarised by the coefficient of variation (CV) across years, and seasonal amplitude as the peak-to-trough range of the transmission rate divided by twice its annual mean. The association between prevalence and $\mathcal{R}$ was assessed with Spearman's rank correlation, and residual spatial autocorrelation with Moran's I (Moran, 1950).

Model adequacy was assessed in two ways rather than by convergence statistics alone. First, in-sample posterior predictive checks compared the observed county-month counts with replicated data drawn from the posterior predictive distribution, using the proportion of observations within their 95% posterior predictive intervals and a posterior predictive p-value for the sum of squared residuals (Gelman et al., 2013). Second, out-of-sample prediction was assessed (Mulati, 2025). The model was fitted to 2015–2022 data and used to predict 2023–2024 (temporal hold-out), refitted with each county left out in turn (leave-one-county-out), and assessed under both forms of hold-out combined. Predictions were compared with observations using the root mean square error (RMSE), the mean absolute error (MAE), Pearson correlation, the coverage of 95% posterior predictive intervals and, for hotspot detection, the area under the ROC curve (AUC), and were benchmarked against a non-spatial SEIR model, a static spatial model, a non-hierarchical Bayesian model and a geostatistical model. Sensitivity was assessed in three ways: structural sensitivity, by refitting with fixed connectivity weights, without seasonal terms, with an exponential rather than Matérn correlation, with linear rather than adaptive temporal adaptation and with alternative priors; parameter sensitivity, by Morris screening followed by Sobol variance decomposition over 15 parameters and 1,000 parameter combinations; and data-quality sensitivity, by introducing 10%, 20% and 30% missing observations, random measurement noise and systematic bias.

The effect of treatment on transmission potential was examined analytically with the normalised forward sensitivity index $\Upsilon_p = (\partial\mathcal{R}/\partial p)(p/\mathcal{R})$, which gives the percentage change in $\mathcal{R}$ produced by a 1% change in parameter p. From Equation (4g), the indices for the treatment initiation rate γ and the treatment rate among AIDS cases τ are

$$\Upsilon_\gamma = -\frac{\gamma}{\gamma+\rho+\mu}, \tag{5a}$$

$$Y_\tau = -\frac{\tau\eta\rho}{(\alpha+\tau+\mu)(\alpha+\tau+\mu+\eta\rho)}. \quad (5b)$$

For changes larger than a few percent the exact ratio was used instead of the linear index. Raising γ to γ' multiplies the reproduction number by a constant factor, and the treatment initiation rate at which a county's reproduction number reaches the threshold of one follows directly:

$$\frac{\mathcal{R}(\gamma')}{\mathcal{R}(\gamma)} = \frac{\gamma+\rho+\mu}{\gamma'+\rho+\mu}, \quad (6)$$

$$\gamma_c = \mathcal{R}(\gamma)(\gamma+\rho+\mu) - (\rho+\mu). \quad (7)$$

These calculations rest on the following assumptions. First, only the treatment parameter is changed; $\beta_i, w_{ij}, \sigma, \rho, \alpha, \eta$ and μ are held at their baseline posterior means, so the index is a local derivative rather than a joint effect. Second, treatment acts only through the duration of infectiousness: γ enters $\mathcal{R}$ only through $1/(\gamma+\rho+\mu)$ in V, and it is assumed not to change contact rates, connectivity or sexual behaviour, so risk compensation is not represented. Third, people on treatment are virally suppressed, do not transmit and do not interrupt treatment (Cohen et al., 2016; Rodger et al., 2019), so a higher γ removes people from the infectious pool completely; where adherence or suppression is incomplete, the calculated effect is an upper bound. Fourth, the susceptible population is evaluated at the disease-free equilibrium, so S_i^0/N_j^0 does not change with γ and $\mathcal{R}$ describes transmission potential rather than short-term incidence. Fifth, rates are constant over time, so the mean time to treatment initiation is $1/\gamma$. Sixth, a given change in γ applies uniformly within each county to which it is applied. Finally, in the global (Morris and Sobol) analysis, γ was not varied alone but sampled jointly with the other 14 parameters by Latin hypercube sampling, so its Sobol index measures the share of output variance attributable to γ rather than the effect of changing treatment alone.

RESULTS

Model Convergence and Fit

All 11 monitored model parameters converged, with Gelman–Rubin statistics $(\hat{R})$ between 1.01 and 1.04, below the 1.05 threshold, and effective sample sizes of 1,156–1,247 for the key transmission and spatial parameters (Mulati, 2025). A stricter threshold of $\hat{R} < 1.01$ was not reached; this is acknowledged as a limitation. The sampler was run with 4 chains of 2,000 iterations each (1,000 warm-up), and 0 divergent transitions were recorded. Residual Moran's I was 0.03 ($p = 0.45$), indicating that the model left no significant spatial autocorrelation in its errors. Table 3 presents the posterior estimates of the key parameters.

Table 3: Posterior Estimates of Key Model Parameters

Parameter	Mean	95% Credible Interval	Interpretation
Baseline transmission rate (β_0)	0.122	0.089–0.156	Average transmission intensity
Spatial variance (σ_θ^2)	0.453	0.321–0.608	Size of unexplained spatial variation
Spatial correlation range (ℓ, km)	85.3	62.1–108.5	Distance over which risk is shared

The posterior range of 85.3 km indicates that unexplained transmission risk is shared between neighbouring counties but decays to near zero between regions hundreds of kilometres apart.

Model Checking and Sensitivity

Table 4 summarises the posterior predictive checks, out-of-sample checks and sensitivity analyses. The model reproduced held-out county data closely (RMSE 0.035, correlation 0.86), with prediction error 22% lower than the best benchmark (the static spatial model) and 48% lower than the non-spatial SEIR model, and within-county temporal fit gave R^2 of at least 0.87 in every county. In sample, 94.7% of observed county-month counts fell within their 95% posterior predictive intervals and the posterior predictive p-value for the sum of squared residuals was 0.38, indicating no systematic misfit. Out of sample, coverage of the 95% predictive intervals exceeded 85% at all horizons but fell short of the nominal 95%, which suggests that the intervals are somewhat

too narrow and that the uncertainty reported below is, if anything, understated. Every structural simplification increased prediction error, most of all fixing the connectivity weights (18%) and removing the seasonal terms (15%), whereas alternative priors changed it by only 5%. Performance remained acceptable with up to 20% of observations missing, and systematic bias degraded it more than random noise. The baseline transmission rate accounted for the largest share of output variance, followed by the spatial variance and range; the treatment rate γ ranked fourth, so estimated transmission potential is moderately sensitive to assumptions about treatment. These checks concern the prediction of county case counts; the stability of individual county reproduction numbers under alternative specifications and county selections was not assessed directly.

Analytically, $|\Upsilon_\gamma| < 1$ for all parameter values (Equation 5a), so a given percentage increase in the treatment initiation rate reduces the reproduction number by a smaller percentage; the reduction approaches proportionality only when treatment initiation is much faster than progression to AIDS and background mortality. By Equation (7), bringing Homa Bay ($\mathcal{R} = 3.2$) to the threshold would require the total rate of leaving untreated chronic infection, $\gamma + \rho + \mu$, to rise about 3.2-fold, with all other parameters unchanged.

Table 4: Model Checking and Sensitivity Results

Check	Result	Interpretation
In-sample posterior predictive check	94.7% of observed county-month counts within 95% posterior predictive intervals; posterior predictive p = 0.38 (sum of squared residuals)	No systematic misfit; in-sample coverage close to nominal
Out-of-sample prediction (temporal and leave-one-county-out)	RMSE 0.035; MAE 0.027; r = 0.86	Lowest error of the five models compared
Benchmark models (RMSE)	Non-spatial SEIR 0.067; static spatial 0.045; non-hierarchical Bayesian 0.052; geostatistical 0.048	Error 22–48% lower than benchmarks
Hotspot detection (AUC)	0.87; 0.78 for emerging hotspots 24 months ahead	Good discrimination
Out-of-sample coverage of 95% predictive intervals	Above 85% at all horizons	Below nominal; intervals somewhat narrow
Within-county temporal fit	$R^2 \geq 0.87$ in every county	Temporal dynamics captured
Residual spatial autocorrelation	Moran's I = 0.03 (p = 0.45)	No residual clustering
Structural alternatives (change in RMSE)	Fixed weights +18%; no seasonal terms +15%; exponential correlation +12%; linear adaptation +8%; alternative priors +5%	Adaptive connectivity and seasonality improve fit; priors have little influence
Missing data (change in RMSE)	10% missing +5%; 20% +12%; 30% +25% (minor convergence problems)	Acceptable up to 20% missing
Data error (change in RMSE)	Random noise (CV 0.3) +18%; systematic bias +22%	Bias more damaging than noise
Parameter sensitivity (Sobol total-effect index)	β_0 0.42; σ_θ 0.25; ℓ 0.19; γ 0.14; ρ 0.11; σ 0.08; μ 0.06	Transmission and spatial parameters dominate; moderate sensitivity to treatment rate
Identifiability (Fisher information eigenvalues)	Leading few eigenvalues above threshold	Several parameters weakly identified
Share of predictive uncertainty	Parameters 45%; spatial process 35%; observation 20%	Parameter uncertainty largest

Note: Results from Mulati (2025), except the in-sample posterior predictive check. RMSE, root mean square error; MAE, mean absolute error; AUC, area under the receiver operating characteristic curve; CV, coefficient of variation.

County Reproduction Numbers

Table 5 presents HIV prevalence and the estimated reproduction number for each county, ordered by $\mathcal{R}$.

Table 5: HIV Prevalence and Reproduction Number by County

County	Region	Adult HIV prevalence (%)	Local reproduction number ($\mathcal{R}$)	95% CrI	Transmission class
Homa Bay	Western	21.0	3.2	2.9–3.5	High
Siaya	Western	20.3	3.1	2.8–3.4	High
Kisumu	Western	18.5	2.8	2.5–3.1	High
Mombasa	Coast	11.4	1.8	1.5–2.1	Medium
Turkana	Northern	6.8	1.5	1.2–1.8	Medium
Nairobi	Central	6.1	1.4	1.2–1.7	Medium
Kwale	Coast	5.7	1.3	1.1–1.6	Medium
Kilifi	Coast	4.5	1.2	1.0–1.5	Medium
Nakuru	Central	4.2	1.2	1.0–1.5	Medium
Kiambu	Central	3.8	1.1	0.9–1.3	Medium
Machakos	Eastern	3.9	1.1	0.9–1.3	Medium
Kitui	Eastern	3.7	1.0	0.8–1.2	Medium
Meru	Eastern	3.4	1.0	0.8–1.2	Medium
Garissa	North Eastern	1.1	0.5	0.3–0.7	Low
Wajir	North Eastern	0.8	0.3	0.2–0.5	Low

Note: $\mathcal{R}$ is the posterior mean county-specific (local) control reproduction number, $\mathcal{R}_i = K_{ii}$ (Eq. 4h), under prevailing treatment coverage; CrI, credible interval. Prevalence is the adult HIV prevalence in the study dataset, taken from the National AIDS Control Council's Kenya HIV Estimates (NACC, 2022) as compiled in Mulati (2025). These are model-based county estimates rather than single-survey estimates and therefore differ from the KENPHIA 2018 figures cited in the Introduction (for example, 21.0% against 19.6% for Homa Bay, 18.5% against 17.5% for Kisumu, and 0.8% for Wajir against less than 0.2% in parts of the former North Eastern region), reflecting differences in source, method and reference period.

As shown in Table 5, transmission potential varied about tenfold, from $\mathcal{R} = 0.3$ in Wajir to $\mathcal{R} = 3.2$ in Homa Bay. Three counties were high transmission, ten medium and two low. Eleven counties had $\mathcal{R}$ above one, while Kitui and Meru lay at the threshold. The ranking of counties by $\mathcal{R}$ closely followed their ranking by prevalence (Spearman $\rho_s = 0.997$, $p < 0.001$). The departures were informative: Turkana's transmission potential (1.5) exceeded that of Nairobi (1.4) despite similar prevalence, and Mombasa (1.8) stood clearly above every other county outside the western region. Because K is non-negative, the system reproduction number satisfies $\mathcal{R}_{\text{sys}} = r(K) \geq \max_i \mathcal{R}_i = 3.2$ (Eq. 4h), so the connected study area as a whole remained above the epidemic threshold. Because $\mathcal{R}_{\text{sys}} \geq \max_i \mathcal{R}_i$ holds in every posterior draw, the posterior distribution of $\mathcal{R}_{\text{sys}}$ lies at or above that of the largest local value: its 2.5th percentile is at least 2.9, the lower credible limit for Homa Bay, and the posterior probability that the connected system is above the threshold is effectively one. Uncertainty matters most for counties near the threshold. The 95% credible interval lay entirely above one for seven counties (Homa Bay, Siaya, Kisumu, Mombasa, Turkana, Nairobi and Kwale) and entirely below one for Garissa and Wajir. For the remaining six (Kilifi, Nakuru, Kiambu, Machakos, Kitui and Meru) the interval reached or crossed one, so whether local transmission in these counties is self-sustaining cannot be determined from the present data.

Transmission Hotspots and Connectivity

Homa Bay, Siaya and Kisumu met every element of the hotspot definition. All three were high transmission, all three border Lake Victoria's Winam Gulf within about one correlation range of each other, and they were joined by the strongest connectivity weights in the network. They formed the only transmission hotspot among the study counties. Table 6 summarises the connectivity weights between selected county pairs over 2015–2024, and Figure 2 shows their trajectories.

Table 6: Connectivity Weights Between Selected County Pairs, 2015–2024

County pair	Mean	Minimum	Maximum	CV (%)	Pattern
Homa Bay–Siaya	0.86	0.84	0.88	1.4	Strong, stable
Kisumu–Homa Bay	0.73	0.71	0.75	1.6	Strong, stable
Nairobi–Mombasa	0.39	0.32	0.45	9.7	Moderate, variable
Garissa–Wajir	0.15	0.14	0.17	5.5	Weak, stable

Note: Table 6 reports posterior means of connectivity weights. Because individual pairwise weights are weakly identified (see the identifiability analysis), 95% credible intervals are not tabulated; the robust features are the ordering and stability of the western cluster.

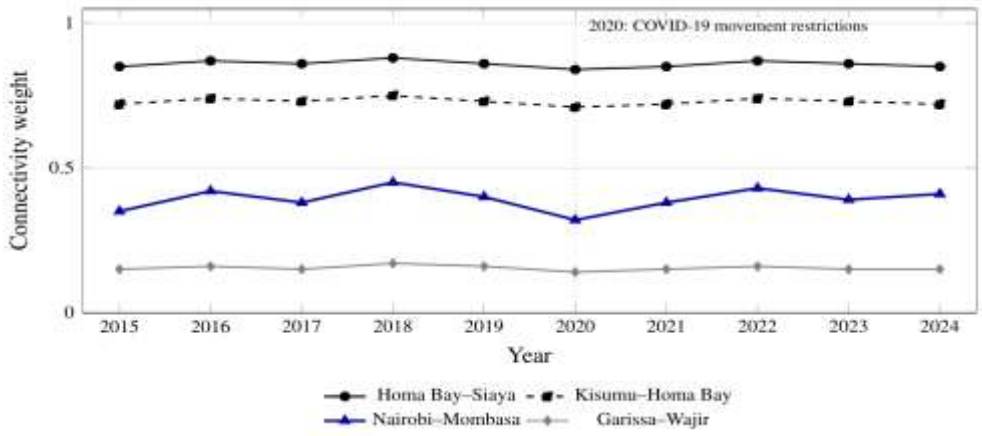

Figure 2: Connectivity Weights Between Selected County Pairs, 2015–2024

Source: Researcher (2026)

Connectivity within the western cluster was strong and almost constant. The Garissa–Wajir link in the northeast was weak and also stable. The Nairobi–Mombasa corridor, which links Kenya's two largest cities along the main transport route, was the most variable of the four pairs. Its weight peaked at 0.45 in 2018 and fell to its lowest value, 0.32, in 2020, a 20% fall from 2019 that coincided with COVID-19 restrictions on movement into and out of both cities (Aluga, 2020). It recovered to 0.43 by 2022.

Seasonal Patterns of Transmission

Table 7 and Figure 3 summarise the seasonal pattern of the full fitted monthly transmission rate by county burden group.

Table 7: Seasonal Variation in the Full Fitted Monthly Transmission Rate by County Burden Group

Burden group	Peak month	Trough month(s)	Peak/trough ratio	Seasonal amplitude (%)
High burden (Western)	October	March	2.3	41
Medium burden	October	March	1.3	12
Low burden (North Eastern)	July	December–February	2.3	41

Note: Peak and trough months are posterior summaries of the full fitted monthly transmission rate. The annual harmonic in Eq. (3b) is one component of the model; the reported months are not the argmax and argmin of that harmonic in isolation. Because a single annual harmonic cannot by itself produce an October peak with a March trough, the peak and trough months reported here are descriptive summaries of the full model output rather than direct estimates from the harmonic. Individual seasonal coefficients are weakly identified, so the peak months and amplitudes are interpreted descriptively rather than as precise estimates.

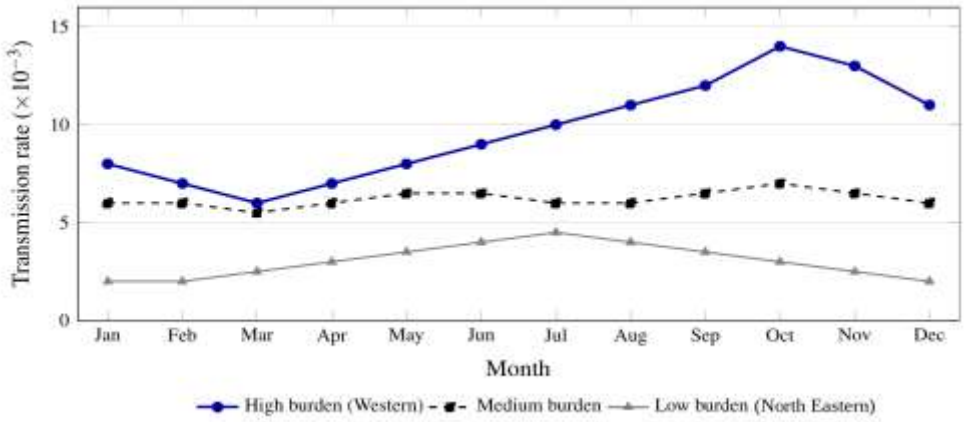

Figure 3: Seasonal Transmission Rate by County Burden Group

Source: Researcher (2026)

In high-burden western counties, the transmission rate rose steadily from a March low to an October peak, a 2.3-fold range. Medium-burden counties showed only a modest October peak. Low-burden north-eastern counties peaked earlier, in July, against a December–February low. Removing the seasonal terms or replacing adaptive connectivity with fixed weights increased prediction error by 15% and 18% respectively (Table 4), suggesting that both features capture systematic structure in the data rather than noise.

DISCUSSION

The findings show that HIV transmission potential in Kenya is concentrated in one place: a single, tightly connected cluster of three Lake Victoria counties with reproduction numbers between 2.8 and 3.2. Outside that cluster, most study counties lie just above the epidemic threshold, and the north-east lies well below it. This pattern is consistent with the 2018 national survey, which recorded the three highest county prevalences in the same area (NASCOP, 2022), and with continental and Kenyan Bayesian maps that place the Lake Victoria basin at the top of the national risk surface (Dwyer-Lindgren et al., 2019; Ngesa et al., 2014; Otiende et al., 2019). The present study adds that this concentration does not appear to be only a legacy of past infection: under prevailing treatment coverage, the model estimates that each infection in these counties is associated with about three further infections, which is consistent with a cluster that remains self-sustaining.

The low reproduction numbers in Garissa and Wajir agree with the near-absence of infection recorded in the North Eastern region (NASCOP, 2022). Values below one suggest that, under current conditions, local chains of transmission are unlikely to sustain themselves, so remaining infections there may be imported or longstanding rather than locally generated. Because these counties remain connected to higher-transmission areas, continued importation is plausible, which argues for surveillance and rapid response rather than withdrawal of services. The close agreement between $\mathcal{R}$ and prevalence ($\rho_s = 0.997$) is expected but should not be over-interpreted. The reproduction number in Equation (4g) is evaluated at the disease-free equilibrium and does not itself depend on the number of people infected; rather, the county transmission rates are estimated from case counts that scale with the size of the infected population, so counties with higher prevalence tend to receive higher estimated transmission rates. The reproduction number adds most where it departs from prevalence. Turkana is the clearest example: its transmission potential exceeds that of counties with similar prevalence, which is consistent with cross-border movement and the mobility of pastoralist and displaced populations, although the present design cannot separate those drivers.

These results should be read as associations at county level rather than as evidence of cause and effect. The model estimates how transmission potential is distributed and how it co-varies with connectivity and season, but it cannot show that connectivity or seasonal factors themselves drive transmission. Feedback between the epidemic and the response is also plausible: counties with a higher recognised burden have attracted more testing and programme investment, and more intensive testing increases the number of infections detected, which may in turn raise estimated transmission in those counties. Conversely, lower testing in northern counties may understate their transmission potential. Unmeasured county factors, such as the size of key populations,

voluntary medical male circumcision coverage and county programme capacity, may also influence both the reported data and the estimated parameters (Wakefield, 2008). Panel designs that exploit changes in programme coverage over time, linkage with phylogenetic data and explicit modelling of testing intensity would help to separate these explanations. The estimated correlation range of about 85 km corresponds approximately to the scale of the Winam Gulf. Combined with the strong and stable connectivity weights between Homa Bay, Siaya and Kisumu, this is consistent with the three counties functioning epidemiologically as one unit. Under Kenya's devolved system, each county plans and budgets its own HIV response, yet the strongest connectivity weights in the network link these counties to one another (Table 6). Gains in one county are therefore likely to be exposed to continuing transmission next door. This finding is consistent with cohort and genomic evidence from neighbouring Uganda that infections in lakeside and inland communities are closely linked (Grabowski et al., 2014; Ratmann et al., 2020), and suggests that joint planning across the cluster is likely to achieve more than three separate county plans.

The Nairobi–Mombasa corridor was the most variable link in the network. Its sharp fall in 2020 and recovery by 2022 are consistent with the timing of the COVID-19 movement restrictions (Aluga, 2020). Because the weights are estimated from the model's own mobility covariates, this agreement shows that the covariates registered the restrictions; it does not independently validate the connectivity estimates. Direct origin-destination data would allow such validation and could be used to update connectivity in near real time (Wesolowski et al., 2012). Mombasa's elevated transmission potential, the highest outside the west, points to the coast end of this corridor as a secondary priority.

The October peak in western counties and the July peak in low-burden north-eastern counties are new findings for Kenya and should be treated as hypotheses about their drivers. Possible explanations in the west include harvest-season income and movement and fishing cycles on the lake; in the north-east, the mid-year peak coincides with the long dry season, when pastoralist movement is greatest. Some of the signal may also reflect reporting and campaign cycles in routine data. Because the seasonal coefficients are weakly identified and have not been validated against independent incidence data, these patterns do not yet justify re-timing programmes. Their present value is as a testable hypothesis that routine monthly data can confirm or reject. Viewed through spatial epidemic theory, these findings are consistent with the proposition that transmission potential depends not only on conditions within a county but also on its connections to others. Viewed through hierarchical Bayesian theory, they illustrate the value of sharing information across connected counties when county data are sparse and uneven. The findings also show the limits of both theories in this setting. The close agreement between $\mathcal{R}$ and prevalence indicates that, where prevalence data dominate the likelihood, the mechanistic structure adds most at the margins, as in Turkana and Mombasa, rather than across the whole ranking. The near-constant western connectivity weights suggest that a fixed-weight model would have been adequate for tightly linked neighbours, and that the adaptive component matters mainly for long-distance corridors such as Nairobi–Mombasa. Finally, because counties are administrative rather than epidemiological units, the finding that three counties function as one unit implies that, in Kenya, the appropriate unit for applying spatial epidemic theory may be a cluster of counties rather than the single county.

Limitations and Directions for Future Research

This study has several limitations. First, it covers 15 purposively selected counties out of 47; neighbouring high-prevalence counties such as Migori and Busia were not modelled, so the true western cluster is probably larger than three counties, and the findings cannot be generalised to unstudied counties or to other countries in East Africa without further evidence. Second, the design is ecological and associational: county aggregation hides sub-county hotspots such as fishing beaches and transport towns (Tanser et al., 2009), associations at county level do not apply to individuals (Wakefield, 2008), and the analysis cannot establish that connectivity or season cause the patterns observed. Third, the study relies entirely on secondary data from government reports and surveys, which may be affected by reporting bias; inputs come from several sources with different methods and dates; and data quality is weaker in rural and northern counties. Because testing intensity differs between counties, part of the variation in estimated transmission may reflect differences in case detection rather than in transmission. Fourth, omitted variables such as the size of key populations, male circumcision coverage, viral suppression, pre-exposure prophylaxis use and county programme capacity were not modelled and may influence both the data and the estimates. Fifth, mobility data were limited, so connectivity weights depend partly

on modelling assumptions about distance decay, and without phylogenetic or contact data the direction of transmission between counties cannot be confirmed. Sixth, seasonal patterns may partly reflect reporting and campaign cycles rather than transmission, and the seasonal coefficients are weakly identified, so the peaks reported here are exploratory. In addition, the annual harmonic in Eq. (3b) cannot by itself generate the October/March pattern; the peak and trough months are therefore descriptive summaries of the full model output rather than direct estimates from the harmonic. Seventh, no data from Uganda, Tanzania or other neighbouring countries were included, so cross-border transmission around Lake Victoria and in the north is not represented. Eighth, the compartmental structure simplifies HIV natural history: the exposed compartment is treated as non-infectious although acute infection is highly infectious, and people on treatment are assumed neither to transmit nor to interrupt treatment. Ninth, the identifiability analysis showed that several parameters are only weakly identified from 15 counties, and out-of-sample coverage of the 95% predictive intervals (above 85%) was below the nominal level, so posterior uncertainty may be understated. Tenth, convergence met the conventional $\hat{R} < 1.05$ criterion but not the stricter $\hat{R} < 1.01$ now often recommended. Eleventh, the sensitivity analyses addressed predictive accuracy; the stability of individual county reproduction numbers, connectivity weights and seasonal parameters under alternative specifications, treatment assumptions and alternative county selections was not tested directly. Leave-one-county-out validation assessed prediction, not parameter stability. The system reproduction number was bounded from below by the largest local value rather than summarised directly from the posterior next-generation matrix. Twelfth, the prior means for the baseline transmission rate, spatial variance and spatial range coincide numerically with the posterior means in Table 3. Although an alternative-prior sensitivity analysis is reported in Table 4, this coincidence means that the priors cannot be regarded as fully independent of the results, and the posterior precision should therefore be interpreted with that caveat in mind. Finally, the reproduction number is a model-derived quantity that depends on the model structure, the priors and prevailing treatment coverage, and should not be compared directly with reproduction numbers from untreated epidemics. Future research should extend the model to all 47 counties and to sub-county units, incorporate direct mobility, testing-intensity and cross-border data, use panel and quasi-experimental designs together with phylogenetic data to examine causal pathways, and validate the seasonal findings against independent incidence data.

CONCLUSION

This study set out to examine the spatial patterns of HIV transmission in Kenya. The findings show that transmission potential is concentrated in a single, tightly connected cluster of Lake Victoria counties—Homa Bay, Siaya and Kisumu—where, under current treatment coverage, each infection is estimated to be associated with about three further infections. Most other study counties lie just above the epidemic threshold, and the north-east lies below it. Because unexplained risk is shared across county lines within about 85 km and the western counties are strongly connected, the western cluster appears to function as a single epidemic unit. The model also suggests a seasonal cycle that differs between high- and low-burden settings, although this finding is exploratory and requires independent validation. Because the design is ecological and limited to 15 counties, these conclusions describe county-level patterns within the study area and should be tested in broader and finer-scale research.

RECOMMENDATIONS

1. The National AIDS Control Council and the Council of Governors should establish a joint HIV planning and coordination mechanism for Homa Bay, Siaya and Kisumu counties through the Lake Region Economic Bloc. At a minimum, the mechanism should include: (a) a standing cluster technical working group of the three county AIDS and STI coordinators, chaired in rotation and supported by the National AIDS Control Council; (b) a single annual cluster plan with shared targets for HIV testing, pre-exposure prophylaxis uptake and viral suppression; (c) a jointly agreed calendar for testing and prevention campaigns; (d) cross-county referral and client-tracking arrangements for fisherfolk and other mobile populations around the Winam Gulf; (e) quarterly joint reviews based on a shared Kenya Health Information System dashboard of cluster indicators; and (f) a matching budget line for cluster activities in each county's annual development plan and budget.
2. The national HIV programme and county governments should allocate prevention resources according to county transmission potential rather than uniformly or by population alone, using a tiered approach

that takes account of posterior uncertainty and is reviewed annually as estimates are updated: Tier 1 ($\mathcal{R} >$ 2: Homa Bay, Siaya and Kisumu), intensive combination prevention and the largest share of prevention resources; Tier 2 ($\mathcal{R}$ 1.3–2, with credible intervals above one: Mombasa, Turkana, Nairobi and Kwale), targeted programmes for key and mobile populations, including the Nairobi–Mombasa transport corridor and Turkana's border areas; Tier 3 ($\mathcal{R}$ 1.0–1.2, with credible intervals reaching or including one: Kilifi, Nakuru, Kiambu, Machakos, Kitui and Meru), sustained routine prevention aimed at bringing transmission below the threshold, together with better data so that their position relative to the threshold can be resolved; and Tier 4 ($\mathcal{R} < 1$: Garissa and Wajir), surveillance with rapid response to imported cases.

3. County health departments, supported by the National AIDS and STI Control Programme, should use monthly Kenya Health Information System data to test whether the modelled seasonal peaks (October in high-burden western counties and July in low-burden north-eastern counties) recur. If they are confirmed, seasonal timing of HIV testing, pre-exposure prophylaxis and linkage-to-care outreach should first be piloted and evaluated in one or two counties before it is built into annual work plans and commodity distribution.
4. The National AIDS Control Council and the Kenya National Bureau of Statistics should develop routine sources of county-to-county mobility data, for example through data-sharing agreements with mobile network operators that comply with the Data Protection Act, 2019 and provide anonymised, quarterly origin-destination flows between counties, so that connectivity between counties can be monitored and updated in near real time.
5. The Ministry of Health should strengthen HIV reporting through the Kenya Health Information System in northern and rural counties by means of quarterly data quality audits, completeness and timeliness targets for facility reports, and routine reporting of testing volumes alongside positive results, so that differences in case detection can be separated from differences in transmission.
6. Future research should extend the analysis to all 47 counties and to sub-county units, incorporate cross-border data from Uganda and Tanzania around Lake Victoria, use panel and phylogenetic data to examine the causal pathways behind the associations reported here, and investigate the drivers of the seasonal patterns identified.

ACKNOWLEDGEMENTS

The authors thank Dr. John Sirengo and Dr. Lucy Chikamai of the Department of Mathematics, Kibabii University, for their supervision and guidance, and the National AIDS Control Council, the Ministry of Health and the Kenya National Bureau of Statistics for the data used in this study.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

The study used aggregated secondary data from the National AIDS Control Council, the Kenya Health Information System, the national population-based HIV surveys (KAIS and KENPHIA) and the Kenya National Bureau of Statistics. The compiled county-month dataset and analysis code are available from the corresponding author on reasonable request.

APPENDIX A: PRIOR SPECIFICATION

The prior distributions were $\beta_{i0} \sim$ Gamma(2,16.4), $\sigma_\theta^2 \sim$ Inverse-Gamma(2.5,0.68) and $\ell \sim$ Inverse-Gamma(2.5,128). Their 95% prior intervals were 0.015–0.340 for the baseline transmission rate, 0.11–1.64 for the spatial variance and 20–308 km for the spatial range.

Table A1 compares the 95% prior intervals with the corresponding posterior credible intervals.

Table A1: Prior Intervals and Posterior Credible Intervals

Parameter	Prior 95% interval	Posterior 95% CrI
Baseline transmission rate (β_{i0})	0.015–0.340	0.089–0.156
Spatial variance (σ_θ^2)	0.11–1.64	0.321–0.608
Spatial range (ℓ, km)	20–308	62.1–108.5

Note: The prior intervals are about five to six times wider than the posterior credible intervals, indicating that the data determined the precision of these estimates. The numerical equality between the prior means and the posterior means in Table 3 is acknowledged in the Limitations.

REFERENCES

1. Aluga, M. A. (2020). Coronavirus disease 2019 (COVID-19) in Kenya: Preparedness, response and transmissibility. *Journal of Microbiology, Immunology and Infection, 53*(5), 671–673.
2. Anderson, R. M., & May, R. M. (1991). *Infectious diseases of humans: Dynamics and control*. Oxford University Press.
3. Anderson, S.-J., Cherutich, P., Kilonzo, N., et al. (2014). Maximising the effect of combination HIV prevention through prioritisation of the people and places in greatest need: A modelling study. *The Lancet, 384*(9939), 249–256.
4. Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data* (2nd ed.). Chapman and Hall/CRC.
5. Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics, 43*(1), 1–20.
6. Camlin, C. S., Akullian, A., Neilands, T. B., et al. (2019). Gendered dimensions of population mobility associated with HIV across three epidemics in rural Eastern Africa. *Health & Place, 57*, 339–351.
7. Carpenter, B., Gelman, A., Hoffman, M. D., et al. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*(1), 1–32.
8. Cohen, M. S., Chen, Y. Q., McCauley, M., et al. (2016). Antiretroviral therapy for the prevention of HIV-1 transmission. *New England Journal of Medicine, 375*(9), 830–839.
9. Cuadros, D. F., Li, J., Branscum, A. J., et al. (2017). Mapping the spatial variability of HIV infection in Sub-Saharan Africa: Effective information for localized HIV prevention and control. *Scientific Reports, 7*, 9093.
10. Diekmann, O., Heesterbeek, J. A. P., & Metz, J. A. J. (1990). On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology, 28*(4), 365–382.
11. Diekmann, O., Heesterbeek, J. A. P., & Roberts, M. G. (2010). The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface, 7*(47), 873–885.
12. Dwyer-Lindgren, L., Cork, M. A., Sligar, A., et al. (2019). Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. *Nature, 570*(7760), 189–193.
13. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC.
14. Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–472.
15. Grabowski, M. K., Lessler, J., Redd, A. D., et al. (2014). The role of viral introductions in sustaining community-based HIV epidemics in rural Uganda: Evidence from spatial clustering, phylogenetics, and egocentric transmission models. *PLoS Medicine, 11*(3), e1001610.
16. Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research, 15*, 1593–1623.
17. Hollingsworth, T. D., Anderson, R. M., & Fraser, C. (2008). HIV-1 transmission, by stage of infection. *Journal of Infectious Diseases, 198*(5), 687–693.
18. Karuri, J., Waiganjo, P., Orwa, D., & Manya, A. (2014). DHIS2: The tool to improve health data demand

and use in Kenya. *Journal of Health Informatics in Developing Countries, 8*(1), 38–60.
19. Keeling, M. J., & Rohani, P. (2008). *Modeling infectious diseases in humans and animals*. Princeton University Press.
20. Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London, Series A, 115*(772), 700–721.
21. Lawson, A. B. (2018). *Bayesian disease mapping: Hierarchical modeling in spatial epidemiology* (3rd ed.). Chapman and Hall/CRC.
22. Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika, 37*(1–2), 17–23.
23. Mulati, O. N. (2025). *Modeling HIV and AIDS transmission dynamics in Kenya using an adaptive spatial hierarchical Bayesian approach* [Unpublished doctoral thesis, Kibabii University].
24. Mulati, O. N. (2026). Empirical validation of an adaptive spatial hierarchical Bayesian SEIR model for HIV transmission dynamics in Kenya. *Iconic Research and Engineering Journals, 9*(7), 2971–2977. https://doi.org/10.64388/IREV9I7-1713792
25. Mulati, O. N., Sirengo, J., & Chikamai, L. (2025a). Development of an adaptive spatial hierarchical Bayesian SEIR model for HIV and AIDS transmission dynamics in Kenya. *Iconic Research and Engineering Journals, 9*(1), 1118–1122.
26. Mulati, O. N., Sirengo, J., & Chikamai, L. (2025b). Theoretical properties of an adaptive spatial hierarchical SEIR model for HIV transmission dynamics. *Iconic Research and Engineering Journals, 9*(2), 45–58.
27. National AIDS and STI Control Programme (NASCOP). (2022). *Kenya Population-based HIV Impact Assessment (KENPHIA) 2018: Final report*. Ministry of Health, Kenya.
28. National AIDS Control Council (NACC). (2020). *Kenya AIDS Strategic Framework II 2020/21–2024/25*. NACC, Kenya.
29. National AIDS Control Council (NACC). (2022). *Kenya HIV estimates report 2022*. NACC, Kenya.
30. Ngesa, O., Mwambi, H., & Achia, T. (2014). Bayesian spatial semi-parametric modeling of HIV variation in Kenya. *PLoS ONE, 9*(7), e103299.
31. Otiende, V. A., Achia, T. N., & Mwambi, H. G. (2019). Bayesian modeling of spatiotemporal patterns of TB-HIV co-infection risk in Kenya. *BMC Infectious Diseases, 19*, 902.
32. Ratmann, O., Kagaayi, J., Hall, M., et al. (2020). Quantifying HIV transmission flow between high-prevalence hotspots and surrounding communities: A population-based study in Rakai, Uganda. *The Lancet HIV, 7*(3), e173–e183.
33. Republic of Kenya. (2006). *HIV and AIDS Prevention and Control Act, No. 14 of 2006*. Government Printer.
34. Republic of Kenya. (2010). *The Constitution of Kenya, 2010*. National Council for Law Reporting.
35. Republic of Kenya. (2012a). *The County Governments Act, No. 17 of 2012*. Government Printer.
36. Republic of Kenya. (2012b). *The Public Finance Management Act, No. 18 of 2012*. Government Printer.
37. Republic of Kenya. (2012c). *The Intergovernmental Relations Act, No. 2 of 2012*. Government Printer.
38. Rodger, A. J., Cambiano, V., Bruun, T., et al. (2019). Risk of HIV transmission through condomless sex in serodifferent gay couples with the HIV-positive partner taking suppressive antiretroviral therapy (PARTNER): Final results of a multicentre, prospective, observational study. *The Lancet, 393*(10189), 2428–2438.
39. Tanser, F., Bärnighausen, T., Cooke, G. S., & Newell, M.-L. (2009). Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. *International Journal of Epidemiology, 38*(4), 1008–1016.
40. van den Driessche, P., & Watmough, J. (2002). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences, 180*(1–2), 29–48.
41. Wakefield, J. (2008). Ecologic studies revisited. *Annual Review of Public Health, 29*, 75–90.
42. Wesolowski, A., Eagle, N., Tatem, A. J., et al. (2012). Quantifying the impact of human mobility on malaria. *Science, 338*(6104), 267–270.