

Comparative Spatial Regression Modeling of HIV Cases in Indonesia in 2023

Andi Mariani ^{1*}, Maftuhul Husnaeni ², Annisa Nurariqah ³, Cantika Lestari ⁴, Muhammad I. Ramadhan ⁵

¹ University of Osaka, Jepang, email: marianiandi@email.com

² Universitas Negeri Makassar, Indonesia, email: husnaeni@email.com

³ Universitas Negeri Makassar, Indonesia, email: nurariqah@email.com

⁴ Universitas Negeri Makassar, Indonesia, email : lestari@email.com

⁵ Universitas Negeri Makassar, Indonesia, email : ramadhan@email.com

Abstract

Reported HIV cases vary substantially across Indonesian provinces and may exhibit spatial dependence that is not adequately represented by conventional regression. This study compared spatial autoregressive (SAR), spatial error (SEM), and General Spatial Model (GSM) specifications for reported HIV cases in 34 Indonesian provincial units in 2023. The response was the annual number of reported HIV cases, and the explanatory variables were population density, the Gini Index, the open unemployment rate, and mean years of schooling. Global Moran's I, Lagrange Multiplier tests, maximum-likelihood estimation, and Akaike information criterion (AIC) were used to assess spatial dependence and compare models. Reported HIV cases exhibited significant positive spatial autocorrelation ($I = 0.61893$, $p < .001$). Robust LM-lag was significant ($p = .026019$), whereas robust LM-error was not significant ($p = .950388$), providing stronger support for a spatial lag specification. SAR had the lowest AIC (594.07), compared with GSM (595.67) and SEM (596.00), and was therefore selected as the preferred model. In the SAR model, population density, the Gini Index, and the open unemployment rate were positively associated with reported HIV cases, while mean years of schooling was negatively associated; all four explanatory variables were statistically significant. The spatial lag parameter was positive but not significant ($\rho = 0.21451$, $p = .10170$). These ecological findings support geographically coordinated HIV surveillance and prevention while requiring cautious interpretation because reported case counts reflect population size, testing, and reporting coverage.

Keywords: *Human Immunodeficiency Virus (HIV); Spatial Regression; Spatial Autoregressive Model; Spatial Autocorrelation; Socioeconomic Factors*

Introduction

Human immunodeficiency virus (HIV) remains a major public health challenge because it attacks the immune system and, without effective treatment, can progress to acquired immunodeficiency syndrome (AIDS). HIV is transmitted primarily through unprotected sexual contact, sharing contaminated injecting equipment, exposure to infected blood, and mother-to-child transmission. Although antiretroviral therapy has substantially improved survival and reduced onward transmission, unequal access to prevention, testing, and treatment continues to sustain the epidemic, particularly among populations facing social and economic disadvantage (World Health Organization [WHO], 2024; Joint United Nations Programme on HIV/AIDS [UNAIDS], 2023).

Indonesia continues to face a substantial HIV burden. The Ministry of Health reported 57,299 HIV cases and 16,410 AIDS cases in 2023, indicating that transmission remains an important national health concern (Kementerian Kesehatan Republik Indonesia [Kemenkes], 2023). However, the reported cases are not evenly distributed across the country. Differences in population concentration, mobility, socioeconomic conditions, educational attainment, health-

<https://doi.org/10.30605/likelihood.1687>

service availability, testing coverage, and reporting practices may produce considerable variation among provinces. Consequently, a national total alone is insufficient for identifying where the burden is concentrated and which contextual factors are associated with provincial differences in reported HIV cases.

The geographical distribution of HIV is relevant because transmission and case detection do not occur independently across administrative boundaries. Population mobility, economic connections, referral networks, and access to health services can link neighboring regions. A province with a high number of reported cases may therefore be surrounded by provinces with similar conditions, while unmeasured regional characteristics may also generate spatially correlated errors. Ignoring these connections violates the independence assumption of conventional regression and may result in inefficient estimates, biased statistical inference, or an incomplete understanding of the geographical pattern (Anselin, 1988; Auchincloss et al., 2012; Elliott & Wartenberg, 2004; LeSage & Pace, 2009).

Spatial regression provides a framework for evaluating these interregional relationships. A spatial autoregressive model (SAR) represents dependence in the response through a spatially lagged dependent variable, whereas a spatial error model (SEM) represents dependence arising from omitted or unobserved factors contained in the error process. The General Spatial Model (GSM), also referred to in this study as a combined spatial lag and spatial error model, includes both mechanisms. Thus, GSM is useful when the number of HIV cases in one province may be associated with values in neighboring provinces and when unobserved influences also follow a spatial pattern (Beale et al., 2010; Elhorst, 2014; Kissling & Carl, 2008; LeSage, 1999; Pramoedyo & Trisilia, 2012; Syahida & Hajarisman, 2021). Because the present study uses cross-sectional data for 2023, GSM is employed to evaluate spatial dependence rather than temporal dynamics.

Four contextual factors are considered in this study: population density, the Gini Index, the open unemployment rate, and mean years of schooling. Population density may be associated with the number of reported HIV cases because densely populated areas generally have more frequent social interaction, greater population mobility, and larger numbers of people potentially requiring testing and treatment. High-density provinces may also contain major urban centers that attract migrants and provide more HIV diagnostic services, increasing both transmission opportunities and case detection. Spatial HIV studies have shown that observed clustering can reflect mobility, settlement patterns, service access, and geographically structured exposure, although the magnitude of these relationships depends on context and on whether analyses use case counts or population-adjusted rates (Dwyer-Lindgren et al., 2019; Solomon et al., 2021; Tanser et al., 2009).

Income inequality, represented by the Gini Index, reflects differences in the distribution of economic resources within a province. Greater inequality may limit access to health information, prevention services, testing, and sustained treatment among disadvantaged groups. It may also intensify social vulnerability and create unequal exposure to structural risks. International evidence indicates that HIV patterns are associated with broader social and economic structures, although the direction and magnitude can differ across settings and epidemic stages (Buot et al., 2014; Drain et al., 2004; Fox, 2010; Poundstone et al., 2004). This provincial association does not imply that every individual living in a high-inequality province has a higher personal risk of infection.

The open unemployment rate is included because unemployment can reduce economic security and restrict access to health services and preventive information. At the population level, prolonged unemployment may be associated with migration, informal employment, and other socioeconomic vulnerabilities that influence access to HIV prevention and care. Area-level health research emphasizes that labor-market conditions and neighborhood socioeconomic environments can shape access to resources and services (Diez Roux, 2001; Krieger, 2001;

Poundstone et al., 2004). However, unemployment should not be treated as a direct cause of HIV infection; rather, it represents a contextual condition that may operate through multiple social, behavioral, and health-service pathways.

Mean years of schooling represents the general educational attainment of the population. Education may improve health literacy, knowledge of transmission routes, awareness of testing, and the ability to access preventive and treatment services. Provinces with higher educational attainment may consequently be better positioned to implement HIV prevention and reduce delayed diagnosis. Educational attainment also reflects broader socioeconomic opportunity and institutional capacity, which are recognized components of population health inequality (Diez Roux, 2001; Krieger, 2001). At the same time, higher educational attainment may coexist with better case-detection systems, so its relationship with reported case counts must be interpreted carefully.

These four explanatory variables represent distinct but interconnected dimensions of the provincial context. Population density describes the concentration of people and potential interaction, the Gini Index reflects the distribution of economic resources, the open unemployment rate indicates labor-market vulnerability, and mean years of schooling captures accumulated educational attainment. Their inclusion in one model makes it possible to examine whether each factor retains an association with reported HIV cases after the other observed provincial characteristics have been considered. This multivariable perspective is important because a bivariate relationship may be misleading when provinces differ simultaneously in urbanization, inequality, employment, and education.

The use of reported case counts also requires careful interpretation. A large number of cases may reflect a greater underlying burden, but it may also reflect a larger population, more active testing, better surveillance, or easier access to diagnostic services. Conversely, a low reported count does not necessarily indicate low transmission when testing and reporting coverage are limited. Therefore, the present analysis evaluates the geographical distribution of reported HIV cases rather than estimating individual infection risk. The findings should be interpreted as ecological associations that can identify regional patterns and generate policy-relevant hypotheses, not as evidence that a provincial characteristic directly causes infection among individuals (Greenland & Morgenstern, 1989).

A spatial perspective adds information that is not available from ordinary regression alone. If neighboring provinces share similar levels of HIV cases because of mobility, service referral, or broader regional conditions, a conventional model may attribute part of this dependence to the observed predictors. Likewise, if omitted contextual factors are spatially clustered, their effects may remain in the residuals. Distinguishing response dependence from error dependence is therefore important for both model specification and policy interpretation (Dormann et al., 2007; Getis, 2007). A spatial lag mechanism may suggest that regional conditions are connected across boundaries, whereas a spatial error mechanism indicates that relevant unobserved processes also follow a geographical pattern. These mechanisms do not by themselves prove direct transmission between provinces, but they help determine whether the observed data are spatially structured.

Several studies have examined the geographical pattern and determinants of HIV/AIDS. Indonesian studies have mapped local risk and service accessibility, applied provincial spatial regression, and combined geographically weighted and spatial autoregressive approaches (Djuraidah et al., 2023; Hanum et al., 2024; Rahmah & Yeni, 2024; Thahar & Sirait, 2023). International studies similarly demonstrate that HIV distributions can form localized clusters and that the observed patterns may vary with population structure, mobility, socioeconomic conditions, and access to services (Dwyer-Lindgren et al., 2019; Solomon et al., 2021; Tanser et al., 2009). Together, these findings show that location is integral to understanding reported HIV patterns.

These studies confirm the importance of location in HIV analysis, but several issues remain open. First, local studies cannot directly describe the national configuration of interprovincial dependence. Second, models that include only a spatial lag or only a spatial error process may not fully represent data in which both mechanisms are plausible. Third, socioeconomic inequality and educational attainment have not always been examined simultaneously with population density and unemployment. Finally, previous applications have not always compared alternative spatial specifications under a common weighting scheme and parsimony criterion.

This study addresses these issues by analyzing reported HIV cases across Indonesian provinces in 2023 using a General Spatial Model and comparing it with SAR and SEM specifications. The number of reported HIV cases is treated as the response variable, while population density, the Gini Index, the open unemployment rate, and mean years of schooling are treated as explanatory variables. The spatial weighting matrix represents the geographical relationship among provinces and enables the analysis to assess whether nearby regions exhibit related case patterns. The analysis focuses on statistical associations at the provincial level; it does not infer individual-level risk or establish causality.

The contribution of this study is threefold. It provides a national provincial-level assessment of the spatial pattern of reported HIV cases, evaluates demographic and socioeconomic factors within a unified spatial regression framework, and compares alternative models to determine whether spatial lag, spatial error, or their combination provides the most appropriate representation of the data. The results are expected to help identify broad regional patterns that can support geographically targeted prevention, testing, and health-service planning while recognizing that policy decisions also require individual-level epidemiological and programmatic evidence.

Accordingly, this study aims to: (1) examine spatial autocorrelation in reported HIV cases and the selected explanatory variables; (2) estimate SAR, SEM, and GSM specifications; (3) compare their goodness of fit and parsimony; and (4) identify provincial contextual factors associated with the distribution of reported HIV cases in Indonesia in 2023. By clearly distinguishing spatial association from causal effects, the study seeks to provide a more methodologically defensible basis for interpreting regional variation in Indonesia's HIV burden.

Method

Research Design and Unit of Analysis

This study used a quantitative ecological cross-sectional design to examine provincial variation in reported HIV cases in Indonesia in 2023. The province was the unit of analysis, and all variables referred to the same observation year. The analytical dataset contained 34 provincial units for which the response and all explanatory variables were available in a consistent geographical classification. Because the analysis used aggregate provincial data, the estimated relationships represent area-level associations and should not be interpreted as individual-level causal effects.

Data Sources and Variables

Secondary data were compiled from official publications of the Ministry of Health of the Republic of Indonesia and Statistics Indonesia (Kemenkes, 2023; Statistics Indonesia, 2024). The response variable was the annual number of reported HIV cases in each province. The explanatory variables were population density, the Gini Index, the open unemployment rate, and mean years of schooling. Records were matched by provincial identifier, checked for consistent names and units, and assembled into one cross-sectional dataset. No individual-level records or personally identifying information were used.

Table A. Operational Definitions of Study Variables

Variable	Operational definition and unit	Source
Y: Reported HIV cases	Annual number of reported HIV cases in each province; cases	Ministry of Health
X1: Population density	Population per square kilometre; persons/km ²	Statistics Indonesia
X2: Gini Index	Provincial income-distribution inequality; index from 0 to 1	Statistics Indonesia
X3: Open unemployment rate	Share of the labour force without work and actively seeking work; percent	Statistics Indonesia
X4: Mean years of schooling	Average completed years of formal schooling; years	Statistics Indonesia

The response was analyzed as the reported number of HIV cases. Accordingly, the coefficients describe associations with case counts at the provincial level. Differences in provincial population size, testing intensity, and reporting coverage remain relevant when interpreting the results. Continuous explanatory variables were retained in their reported units so that the estimates could be related to the original official statistics.

Spatial Weight Matrix

Spatial relationships among provinces were represented by an $n \times n$ spatial weight matrix W . The diagonal elements were set to zero because a province was not defined as its own neighbor. A nonzero off-diagonal element indicated that two provinces were classified as neighbors in the provincial spatial dataset; otherwise, the element was zero. The matrix was row-standardized so that the weights in each row summed to one. Under this standardization, Wy represents the weighted mean of reported HIV cases in the provinces connected to province i . The same matrix was used in Moran's I tests and in all spatial regression specifications to maintain comparability (Bivand et al., 2013).

Model Specification

An ordinary least squares model was first estimated as a nonspatial baseline. The general spatial specification used in this study allows dependence in both the response and the disturbance process and is written as follows:

$$y = \rho Wy + X\beta + u$$

$$u = \lambda Wu + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2 I)$$

where y is the $n \times 1$ vector of reported HIV cases; X is the $n \times (k + 1)$ matrix containing an intercept and the four explanatory variables; β is the coefficient vector; W is the row-standardized spatial weight matrix; ρ is the spatial-lag parameter; λ is the spatial-error parameter; u is the spatially correlated disturbance; and ε is an independently distributed error vector with variance σ^2 . Setting $\lambda = 0$ produces the SAR model, setting $\rho = 0$ produces the SEM model, and estimating both parameters produces the GSM specification.

Diagnostic and Spatial Dependence Tests

Before estimating the spatial models, the baseline regression was assessed for residual normality using the Shapiro-Wilk test, multicollinearity using the variance inflation factor, and heteroskedasticity using the Breusch-Pagan test. A VIF below 10 was treated as evidence that severe multicollinearity was absent. Because the Breusch-Pagan test evaluates nonconstant residual variance, a significant result was interpreted as heteroskedasticity and considered when evaluating the reliability of conventional standard errors; it was not used alone as evidence that a geographically weighted model was required.

Global Moran's I was used to evaluate whether the response and explanatory variables exhibited spatial autocorrelation. The null hypothesis stated that the observed values were spatially random with respect to W , whereas the alternative indicated spatial autocorrelation. Statistical significance was evaluated at $\alpha = .05$. Moran's I values greater than their expected value indicate positive spatial association, while values below the expected value indicate negative spatial association (Anselin, 1995; Getis, 2007). Lagrange Multiplier tests were applied to the OLS residuals to guide spatial model specification. LM-lag tested dependence in the spatially lagged response, and LM-error tested dependence in the disturbance process. When the conventional LM tests were both significant, their robust counterparts were examined to determine which form of dependence remained after controlling for the other. A joint SARMA test was also considered as evidence that lag and error dependence might need to be modeled simultaneously (Anselin et al., 1996).

Parameter Estimation and Model Selection

The SAR, SEM, and GSM parameters were estimated by maximum likelihood. Coefficient significance was assessed using z statistics and two-sided p -values at $\alpha = .05$. A coefficient was interpreted as an adjusted ecological association after controlling for the other observed provincial characteristics. For models containing a spatially lagged response, the regression coefficients were not treated as simple marginal effects because changes may propagate through the spatial multiplier (LeSage & Pace, 2009). Consequently, the signs and statistical significance of the estimates were reported, while causal and individual-level interpretations were avoided. Model comparison used the Akaike information criterion as the primary parsimony-adjusted criterion, with a smaller AIC indicating a preferable balance between fit and complexity (Burnham & Anderson, 2002). Log-likelihood and Nagelkerke pseudo-R-squared were reported as supplementary measures. These criteria were evaluated jointly with the LM diagnostics so that a more complex model was not selected solely because it produced a marginally larger log-likelihood or pseudo-R-squared.

Analytical Procedure and Software

The analysis proceeded through data integration and verification, descriptive assessment, OLS estimation, diagnostic testing, construction and standardization of W , Moran's I testing, LM and robust LM testing, estimation of SAR, SEM, and GSM, model comparison, and substantive interpretation. The analyses were conducted in R through RStudio using spatial-data and spatial-regression routines consistent with the *spdep* and *spatialreg* packages. The analytical sequence is summarized in Figure 1.

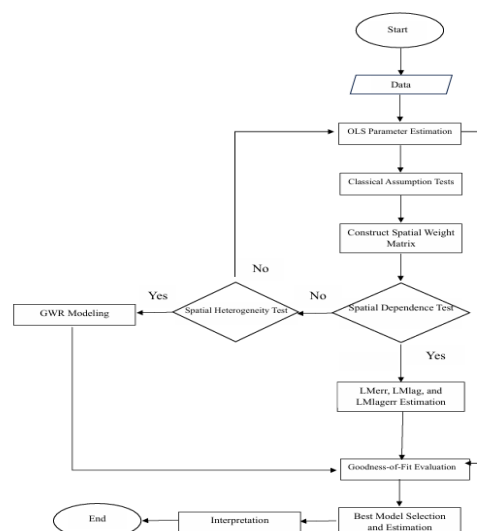


Figure 1. Spatial Regression Analysis Procedure

Results and Discussion

Preliminary Diagnostics

The baseline regression diagnostics are summarized in Table 1. The Shapiro-Wilk test produced a p-value of 0.1874, so the null hypothesis of residual normality was not rejected at $\alpha = .05$. The Durbin-Watson test was also nonsignificant ($p = .2501$). Nevertheless, the Durbin-Watson test is not a substitute for a spatial dependence test in cross-sectional regional data; Moran's I and the Lagrange Multiplier tests provide the more relevant evidence for spatial structure.

Table 1. Baseline Regression Diagnostics

Diagnostic	Statistic	df	p-value	Interpretation
Shapiro-Wilk	-	-	0.1874	Residual normality was not rejected
Durbin-Watson	-	-	0.2501	No conventional serial correlation was detected
Breusch-Pagan	14.135	4	0.006876	Homoscedasticity was rejected

The Breusch-Pagan statistic was 14.135 with $p = .006876$, indicating nonconstant residual variance. This finding requires caution when interpreting conventional standard errors and may reflect differences in provincial scale or reporting. It does not, by itself, establish spatially varying coefficients or justify a geographically weighted regression. The spatial model assessment therefore continued using Moran's I and LM tests, while heteroskedasticity was retained as a limitation of the inference.

Table 2. Variance Inflation Factors

Code	Variable	VIF
X1	Population density	1.439435
X2	Gini Index	1.153079
X3	Open unemployment rate	1.481697
X4	Mean years of schooling	1.689439

All VIF values were below 2, ranging from 1.153079 to 1.689439. Thus, the four explanatory variables did not exhibit problematic multicollinearity, and their simultaneous inclusion did not create substantial linear redundancy.

Global Spatial Autocorrelation

Table 3. Global Moran's I Results

Variable	Moran's I	p-value	Decision
Reported HIV cases (Y)	0.61893133	0.0001452	Significant positive
Population density (X1)	0.09806668	0.2311	Not significant
Gini Index (X2)	0.44307917	0.004043	Significant positive
Open unemployment rate (X3)	0.27236170	0.04441	Significant positive
Mean years of schooling (X4)	-0.13383118	0.7092	Not significant

Reported HIV cases showed strong positive spatial autocorrelation ($I = 0.61893$, $p = .0001452$). Provinces with high reported case counts therefore tended to be spatially connected to provinces with similarly high counts, and the same pattern applied to lower-count areas. The Gini Index ($I = 0.44308$, $p = .004043$) and open unemployment rate ($I = 0.27236$, $p = .04441$) also displayed significant positive spatial clustering. In contrast, population density was not spatially autocorrelated ($p = .2311$), and the negative Moran's I for mean years of schooling was not significant ($p = .7092$). These results correct the earlier interpretation that every positive Moran's I value was automatically significant.

The strong clustering of the response supports the use of spatial regression rather than relying exclusively on OLS. Substantively, clustering may reflect interprovincial mobility, shared service systems, common socioeconomic environments, or similarities in testing and reporting. Moran's I identifies geographical structure but does not determine which mechanism produced it.

Spatial Dependence Tests

Table 4. Lagrange Multiplier Tests

Test	Statistic	df	p-value
LM error	4.0273021	1	0.044769
LM lag	8.9782187	1	0.002732
Robust LM error	0.0038713	1	0.950388
Robust LM lag	4.9547879	1	0.026019
Joint SARMA	8.9820900	2	0.011209

The conventional LM-error and LM-lag tests were significant. When the two forms of dependence were controlled for simultaneously, robust LM-lag remained significant ($p = .026019$), whereas robust LM-error was not significant ($p = .950388$). This pattern gives stronger support to a spatial lag specification than to a spatial error specification. The joint SARMA test was also significant ($p = .011209$), so the GSM was estimated as a comparison; however, this joint diagnostic did not by itself establish that GSM would provide the best final fit.

Spatial Regression Estimates

Table 5. Spatial Autoregressive Model Estimates

Parameter	Estimate	Std. error	z	p-value
Intercept	5658.50	3489.70	1.6215	0.10491
X1	0.23390	0.09026	2.5914	0.00956
X2	12399.00	5681.90	2.1821	0.02910
X3	771.35	178.86	4.3127	0.00001613
X4	-1376.60	341.42	-4.0319	0.00005532
Rho	0.21451	0.13107	1.6366	0.10170

In the SAR model, population density, the Gini Index, and the open unemployment rate were positively associated with reported HIV cases, whereas mean years of schooling was negatively associated with reported cases. All four explanatory variables were statistically significant at $\alpha = .05$. The spatial lag parameter was positive ($\rho = 0.21451$) but was not individually significant ($p = .10170$). The nonsignificant rho does not erase the earlier evidence of clustering, but it indicates that, conditional on the included variables and W , the remaining endogenous response dependence was estimated imprecisely.

The coefficient scales should be read in the original measurement units and as conditional spatial-model estimates rather than simple causal effects. A one-unit increase in population density was associated with an estimated increase of 0.23390 reported cases, while a one-percentage-point increase in the open unemployment rate was associated with 771.35 additional reported cases, holding the other variables constant. Because the Gini Index ranges from 0 to 1, a 0.01 increase corresponds to an estimated 123.99 additional reported cases, not 12,399 cases. One additional year in mean years of schooling was associated with 1,376.60 fewer reported cases. In a spatial lag model, total changes can also include feedback through neighboring regions; therefore, these coefficients primarily establish direction and conditional association.

Table 6. Spatial Error Model Estimates

Parameter	Estimate	Std. error	z	p-value
Intercept	4492.60	3399.30	1.3216	0.18629
X1	0.26645	0.08808	3.0253	0.00248
X2	16213.00	5523.50	2.9353	0.00333
X3	862.27	172.40	5.0016	0.0000005684
X4	-1412.20	340.60	-4.1481	0.0000335
Lambda	0.17793	0.16817	1.0580	0.29004

The SEM produced the same coefficient directions as the SAR model, and all four explanatory variables were significant. However, the spatial error parameter was not significant ($\lambda = 0.17793$, $p = .29004$). This result is consistent with the nonsignificant robust LM-error

test and suggests that, after the observed predictors were included, residual dependence operating exclusively through the error process was not strongly supported.

Table 7. General Spatial Model Estimates

Parameter	Estimate	Std. error	z	p-value
Intercept	6154.20	3552.70	1.7323	0.08323
X1	0.20777	0.09026	2.3018	0.02135
X2	10482.00	6191.90	1.6928	0.09049
X3	732.09	203.28	3.6014	0.00032
X4	-1348.00	360.53	-3.7388	0.00018
Rho	0.30663	0.19453	1.5763	0.11496
Lambda	-0.19751	0.27608	-0.7154	0.47435

The GSM simultaneously estimated response-lag and error dependence. Population density, the open unemployment rate, and mean years of schooling remained significant. The Gini Index was positive but was not significant at the 5% level ($p = .09049$). The corrected statistic for population density was $z = 2.3018$ with $p = .02135$, calculated from its estimate and standard error. Neither rho ($p = .11496$) nor lambda ($p = .47435$) was significant. Estimating both spatial parameters therefore increased model complexity without providing evidence that both mechanisms were required.

Model Comparison

Table 8. Goodness-of-Fit Comparison

Model	Log-likelihood	AIC	Pseudo-R2	Delta AIC
SAR	-290.0360	594.07	0.66413	0.00
SEM	-291.0004	596.00	0.64453	1.93
GSM	-289.8347	595.67	0.66809	1.60

GSM had the largest log-likelihood and pseudo-R-squared, but these improvements were small. Its pseudo-R-squared exceeded that of SAR by only 0.00396. After accounting for model complexity, SAR had the lowest AIC (594.07), followed by GSM (595.67) and SEM (596.00). The AIC differences relative to SAR were 1.60 for GSM and 1.93 for SEM, indicating that the models were relatively close, but SAR provided the best parsimony-adjusted fit. This conclusion is also consistent with the significant robust LM-lag and nonsignificant robust LM-error tests.

Accordingly, SAR was retained as the preferred model for the available data. GSM remains informative as a sensitivity comparison, but it cannot be described as the best model solely because it has the largest log-likelihood and pseudo-R-squared. The nonsignificance of both GSM spatial parameters further weakens such a claim.

Substantive Discussion

The positive association between population density and reported HIV cases is consistent with the concentration of population, mobility, testing facilities, and referral services in major urban provinces. This result should not be interpreted to mean that density mechanically causes infection. Higher-density provinces may identify more cases because they contain larger populations and stronger surveillance systems. Future analyses should therefore compare counts with population-adjusted notification rates and, where possible, include testing coverage.

The positive Gini coefficient indicates that greater provincial income inequality was associated with higher reported case counts in the SAR and SEM specifications. Structural inequality may influence access to prevention, testing, and continuous treatment, although ecological relationships between wealth and HIV can vary across settings and epidemic stages (Drain et al., 2004; Fox, 2010). The loss of 5% significance in GSM also shows that this relationship is sensitive to model specification. Consequently, the direction is policy-relevant, but the magnitude should not be treated as a stable causal effect.

The positive association with the open unemployment rate supports the broader view that HIV disparities are embedded in social and economic conditions rather than being explained only by individual behavior. Employment insecurity can affect mobility, access to services, and continuity of care. Research on distal sociological determinants likewise emphasizes that place-based socioeconomic environments contribute to HIV disparities through interconnected pathways (Buot et al., 2014). The present provincial analysis cannot identify these pathways directly, but it indicates where integrated social and health interventions may be warranted.

Mean years of schooling had a negative and statistically significant coefficient in all three spatial models. Education may improve health literacy, knowledge of prevention, timely testing, and engagement with health services. This result is consistent with the proposed protective role of educational attainment, but it remains an area-level relationship. It does not establish that additional schooling alone would reduce infection, and it may partly represent broader differences in public services and institutional capacity among provinces.

Policy Implications and Study Limitations

The results support geographically targeted HIV programs that integrate epidemiological surveillance with socioeconomic information. Provinces within high-case clusters may benefit from coordinated testing, referral, prevention, and treatment-continuity strategies rather than isolated interventions. The associations with unemployment, inequality, and education also suggest that HIV policy should be linked with social protection, health literacy, stigma reduction, and equitable access to services. These implications are intended to guide prioritization and do not identify individual people or communities as sources of transmission.

Several limitations should be considered. The cross-sectional design cannot establish temporal order or causality. Provincial aggregation creates a risk of ecological fallacy and masks within-province variation (Greenland & Morgenstern, 1989). Reported case counts depend on population size, testing activity, service access, and completeness of surveillance. The analysis contains only 34 spatial units, which limits statistical power, particularly for a two-parameter GSM. Results may also depend on the definition of the spatial weight matrix. In addition, the significant Breusch-Pagan test indicates heteroskedasticity, direct, indirect, and total impacts were not available from the summarized output, and a post-estimation residual Moran's I result was not reported. These limitations should be addressed through population-adjusted rates or spatial count models (Cameron & Trivedi, 2013), alternative weight matrices, residual spatial diagnostics, finer geographical resolution, and multi-year data in future research.

Conclusion

This study compared SAR, SEM, and GSM specifications to analyze provincial variation in reported HIV cases in Indonesia in 2023. The significant positive Moran's I for reported HIV cases ($I = 0.61893$, $p < .001$) confirmed that the response was spatially clustered. The conventional LM tests indicated both lag and error dependence, but only robust LM-lag remained significant after controlling for the competing process. This result provided stronger diagnostic support for SAR. Model comparison reached the same conclusion: SAR had the lowest AIC (594.07), followed by GSM (595.67) and SEM (596.00). Although GSM produced a slightly larger log-likelihood and pseudo-R-squared, the improvement was insufficient to offset its additional complexity, and neither of its spatial parameters was significant. Therefore, GSM was retained as a useful comparison rather than selected as the final model.

In the preferred SAR specification, population density, the Gini Index, and the open unemployment rate were positively associated with reported HIV cases, whereas mean years of schooling was negatively associated. These relationships were statistically significant after the other observed provincial characteristics were considered. The SAR spatial lag parameter was

positive but not significant ($\rho = 0.21451$, $p = .10170$), so the results should not be interpreted as evidence that changes in one province directly caused changes in neighboring provinces. Instead, they show that provincial HIV case patterns are associated with broader demographic, economic, and educational conditions.

The findings support coordinated HIV surveillance, testing, referral, prevention, and treatment planning in geographically connected high-burden areas, together with policies that improve health literacy and equitable access to services. However, the cross-sectional ecological design, the use of reported case counts, the limited number of spatial units, possible underreporting, heteroskedasticity, and sensitivity to the spatial weight matrix restrict causal interpretation. Future research should use population-adjusted rates or spatial count models, finer geographical units, alternative weight matrices, multi-year data, testing-coverage indicators, and direct, indirect, and total spatial impacts.

Acknowledgment

References

- Anselin, L. (1988). *Spatial econometrics: Methods and models*. Kluwer Academic Publishers.
- Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis*, 27(2), 93-115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1), 77-104. [https://doi.org/10.1016/0166-0462\(95\)02111-6](https://doi.org/10.1016/0166-0462(95)02111-6)
- Auchincloss, A. H., Gebreab, S. Y., Mair, C., & Diez Roux, A. V. (2012). A review of spatial methods in epidemiology, 2000-2010. *Annual Review of Public Health*, 33(1), 107-122. <https://doi.org/10.1146/annurev-publhealth-031811-124655>
- Beale, C. M., Lennon, J. J., Yearsley, J. M., Brewer, M. J., & Elston, D. A. (2010). Regression analysis of spatial data. *Ecology Letters*, 13(2), 246-264. <https://doi.org/10.1111/j.1461-0248.2009.01422.x>
- Bivand, R. S., Pebesma, E., & Gómez-Rubio, V. (2013). *Applied spatial data analysis with R* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-4614-7618-4>
- Buot, M. L. G., Docena, J. P., Ratemo, B. K., Bittner, M. J., Burlew, J. T., Nuritdinov, A. R., & Robbins, J. R. (2014). Beyond race and place: Distal sociological determinants of HIV disparities. *PLOS ONE*, 9(4), e91711. <https://doi.org/10.1371/journal.pone.0091711>
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139013567>
- Diez Roux, A. V. (2001). Investigating neighborhood and area effects on health. *American Journal of Public Health*, 91(11), 1783-1789. <https://doi.org/10.2105/AJPH.91.11.1783>
- Djuraidah, A., Anisa, R., Ristiyanti Tarida, A., Alwi Aliu, M., Septemberini, C., Putri Astrini, Y. P., & Tasya Meilania, G. (2023). Pemodelan mixed geographically weighted regression-spatial autoregressive (MGWR-SAR) pada kasus HIV di Indonesia. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 15(2), 65-76. <https://doi.org/10.34123/jurnalasks.v15i2.608>

- Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kühn, I., Ohlemüller, R., Peres-Neto, P. R., Reineking, B., Schröder, B., Schurr, F. M., & Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, 30(5), 609-628. <https://doi.org/10.1111/j.2007.0906-7590.05171.x>
- Drain, P. K., Smith, J. S., Hughes, J. P., Halperin, D. T., & Holmes, K. K. (2004). Correlates of national HIV seroprevalence: An ecologic analysis of 122 developing countries. *Journal of Acquired Immune Deficiency Syndromes*, 35(4), 407-420. <https://doi.org/10.1097/00126334-200404010-00011>
- Dwyer-Lindgren, L., Cork, M. A., Sligar, A., Steuben, K. M., Wilson, K. F., Provost, N. R., Mayala, B. K., VanderHeide, J. D., Collison, M. L., Hall, J. B., Biehl, M. H., Carter, A., Frank, T., Douwes-Schultz, D., Burstein, R., Casey, D. C., Deshpande, A., Earl, L., El Bcheraoui, C., ... Hay, S. I. (2019). Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. *Nature*, 570(7760), 189-193. <https://doi.org/10.1038/s41586-019-1200-9>
- Elhorst, J. P. (2014). *Spatial econometrics: From cross-sectional data to spatial panels*. Springer. <https://doi.org/10.1007/978-3-642-40340-8>
- Elliott, P., & Wartenberg, D. (2004). Spatial epidemiology: Current approaches and future challenges. *Environmental Health Perspectives*, 112(9), 998-1006. <https://doi.org/10.1289/ehp.6735>
- Fox, A. M. (2010). The social determinants of HIV serostatus in sub-Saharan Africa: An inverse relationship between poverty and HIV? *Public Health Reports*, 125(Suppl. 4), 16-24. <https://doi.org/10.1177/00333549101250S405>
- Getis, A. (2007). Reflections on spatial autocorrelation. *Regional Science and Urban Economics*, 37(4), 491-496. <https://doi.org/10.1016/j.regsciurbeco.2007.04.005>
- Greenland, S., & Morgenstern, H. (1989). Ecological bias, confounding, and effect modification. *International Journal of Epidemiology*, 18(1), 269-274. <https://doi.org/10.1093/ije/18.1.269>
- Hanum, H., Yunita Yana, R. D., Andriani, Y., & Setyawan, D. (2024). Spatial regression analysis on factors influencing number of HIV/AIDS cases in Indonesia. *International Journal of Research and Scientific Innovation*, 11(15), 730-738. <https://doi.org/10.51244/ijrsi.2024.11150052p>
- Joint United Nations Programme on HIV/AIDS. (2023). *The path that ends AIDS: UNAIDS global AIDS update 2023*.
- Kementerian Kesehatan Republik Indonesia. (2023). *Laporan kasus HIV dan AIDS di Indonesia tahun 2023*.
- Kissling, W. D., & Carl, G. (2008). Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, 17(1), 59-71. <https://doi.org/10.1111/j.1466-8238.2007.00334.x>
- Krieger, N. (2001). Theories for social epidemiology in the 21st century: An ecosocial perspective. *International Journal of Epidemiology*, 30(4), 668-677. <https://doi.org/10.1093/ije/30.4.668>
- LeSage, J. P. (1999). *The theory and practice of spatial econometrics*. University of Toledo.
- LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. CRC Press. <https://doi.org/10.1201/9781420064254>

- Poundstone, K. E., Strathdee, S. A., & Celentano, D. D. (2004). The social epidemiology of human immunodeficiency virus/acquired immunodeficiency syndrome. *Epidemiologic Reviews*, 26(1), 22-35. <https://doi.org/10.1093/epirev/mxh005>
- Pramoedyo, H., & Trisilia, M. (2012). General spatial models (GSM) approach on baby infant mortality data. *Natural-B*, 1(3), 240-247. <https://doi.org/10.21776/ub.natural-b.2012.001.03.8>
- Rahmah, M., & Yeni. (2024). Analisis spasial kasus HIV/AIDS, aksesibilitas pelayanan kesehatan dan tempat berisiko penularan HIV/AIDS di Kota Palembang tahun 2022. *Media Publikasi Promosi Kesehatan Indonesia*, 7(2), 480-491. <https://doi.org/10.56338/mppki.v7i3.4915>
- Solomon, M., Furuya-Kanamori, L., & Wangdi, K. (2021). Spatial analysis of HIV infection and associated risk factors in Botswana. *International Journal of Environmental Research and Public Health*, 18(7), 3424. <https://doi.org/10.3390/ijerph18073424>
- Statistics Indonesia. (2024). *Statistical yearbook of Indonesia 2024*.
- Syahida, R. S., & Hajarisman, N. (2021). Penerapan general spatial model (GSM) untuk menentukan faktor-faktor penyebab tuberkulosis di Jawa Barat. *Prosiding Statistika*, 7(1).
- Tanser, F., Barnighausen, 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. <https://doi.org/10.1093/ije/dyp148>
- Thahar, S. F., & Sirait, T. (2023). Analisis spasial variabel-variabel yang memengaruhi jumlah kasus baru HIV/AIDS di Provinsi Jawa Timur tahun 2021. *Seminar Nasional Official Statistics*, 2023(1), 287-296. <https://doi.org/10.34123/semnasoffstat.v2023i1.1602>
- World Health Organization. (2024). *HIV and AIDS*.