Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 10, Number 1, 2026
DOI: 10.30829/zero.v10i1.28500
Page: 132-143

Spatial Heterogeneity of Tuberculosis Incidence Using Geographically Weighted Negative Binomial Regression (GWNBR) in Indonesia

[1]Riry Sriningsih
Department of Mathematics, Universitas Negeri Padang, Padang, 25171, Indonesia

[2]Mohammad Soleh
Department of Mathematics, Universitas Islam Negeri Sultan Syarif Kasim, Riau, 28293, Indonesia

[3]Muhammad Subhan
Department of Mathematics, Universitas Negeri Padang, Padang, 25171, Indonesia

[4]Reni Prima Gusty
Medical-Surgical Nursing, Universitas Andalas, Padang, 25163, Indonesia

Article Info

Article history:

Accepted 25 March 2026

Keywords:

Cross Validation;
GWNBR;
MLE;
Spatial;
Tuberculosis.

ABSTRACT

Poisson regression is widely used for count data but relies on the equidispersion assumption, which is often violated in epidemiological data due to overdispersion. Negative Binomial Regression (NBR) addresses this issue by introducing a dispersion parameter. However, both models assume spatial homogeneity of parameters. This study applies Geographically Weighted Negative Binomial Regression (GWNBR) to analyze tuberculosis (TB) cases across 38 provinces in Indonesia in 2024. The response variable is the number of TB cases, with predictors including population density, smoking prevalence (age ≥15), poverty rate, and number of hospitals. Overdispersion was confirmed (deviance/df = 12,020), justifying the use of NBR. Model comparison shows that GWNBR provides improved fit relative to global models, with lower AIC than the NBR model (716.45 vs 732.29). Spatial heterogeneity was confirmed by the Breusch–Pagan test (BP = 21.011; $p < 0.01$). Provinces exhibit distinct patterns of significant determinants; for example, in West Sumatra, poverty and smoking show strong positive local effects, while in several eastern provinces smoking is not significant. These findings highlight the importance of spatially adaptive TB control policies rather than uniform national strategies.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Riry Sriningsih,
Department of Mathematics
Universitas Negeri Padang
Email: rirysriningsih@fmipa.unp.ac.id

1. INTRODUCTION

Tuberculosis (TB) remains a major global health challenge, with Indonesia consistently ranked among the highest-burden countries [1]–[4]. Beyond morbidity and mortality, TB imposes substantial socio-economic losses, particularly among the productive-age population [5]. Achieving the End TB elimination target by 2030 therefore requires not only intensified interventions but also analytically robust and geographically sensitive evidence to guide resource allocation [1], [2], [4].

Methodologically, TB incidence is a count data that frequently exhibits overdispersion, rendering the Poisson equidispersion assumption inappropriate. Negative Binomial regression is widely recommended under such conditions [6], [7]. However, both Poisson and Negative Binomial models are inherently global, assuming spatial stationarity in parameter estimates [8]–[14]. In a country as geographically and socio-demographically heterogeneous as Indonesia, this assumption risks producing oversimplified and potentially misleading policy conclusions.

Spatial epidemiology consistently demonstrates that relationships between health determinants and outcomes are geographically non-stationary. Geographically Weighted Regression (GWR) and its count-data extensions enable local parameter estimation, revealing spatial heterogeneity masked by global models [15]–[19]. Empirical studies in high-burden settings confirm significant spatial variation in TB risk structures [20]–[22]. Nevertheless, Indonesian studies largely remain confined to descriptive mapping or global regression frameworks, with limited integration of overdispersion and spatial non-stationarity within a unified national-scale model [23].

This study addresses that critical gap by applying Geographically Weighted Negative Binomial Regression (GWNBR) to all 38 provinces of Indonesia. A sequential modeling strategy—Poisson estimation, overdispersion testing, Negative Binomial fitting, spatial heterogeneity testing, and final GWNBR implementation—ensures methodological rigor. By jointly accommodating overdispersion and spatial heterogeneity, this research provides more defensible inference and geographically differentiated policy insights to support adaptive TB control strategies.

2. RESEARCH METHOD

2.1 Multicollinearity Detection

Multicollinearity is a condition where there is a strong linear relationship between predictor variables in a regression model. The presence of multicollinearity can lead to unstable parameter estimates and increase the variance of the estimator, making the estimation results and model interpretation less reliable [24]. One commonly used method for detecting multicollinearity is the Variance Inflation Factor (VIF). A VIF value exceeding 10 indicates strong multicollinearity among the predictor variables [25].

2.2 Poisson Regression

Poisson regression states that the mean and variance of the count response are equal. In practical applications, count data often exhibits overdispersion, characterized by variability beyond the mean, making conventional Poisson models inadequate [12], [13]. Overdispersion can be evaluated by analyzing the deviations and Pearson's chi-square statistic relative to their respective degrees of freedom [26]. Values exceeding one indicate a violation of the equidispersion assumption, which means alternative models need to be explored, allowing negative binomial regression to be used when significant overdispersion is observed, as it allows the variance to exceed the mean thru an additional dispersion parameter [13].

2.3 Negative Binomial Regression

Negative binomial regression is an extension of Poisson regression within the Generalized Linear Model (GLM) framework, designed to handle count data exhibiting overdispersion [6], [27]. This model introduces an additional dispersion parameter, so the variance of the response does not have to equal the mean, as assumed in Poisson regression [28]. Recent developments show that negative binomial regression remains relevant and widely used as the basis for developing modern count regression models for complex and heterogeneous data [13], [14].

The general form of the global Negative Binomial model is:

$$Y_i \sim NB(\mu_i, \theta)$$
$$\log(\mu_i) = \beta_0 + \sum_{k=1}^{p} \beta_k X_{ik} \quad (1)$$

where Y_i is the number of TB cases in province i, μ_i is the expected value, and θ is the dispersion parameter.

The significance test for the parameters of the negative binomial regression model simultaneously uses the deviance test as follows [26].

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_p = 0$$
$$H_1: \text{there is at least one } \beta_k \neq 0;\ k = 1, \ldots, p \quad (2)$$

aStatistic test:

$$D\left(\hat{\beta}\right) = -2\ln\left(\frac{L\left(\hat{\omega}\right)}{L\left(\hat{\Omega}\right)}\right) = 2\left(\ln L\left(\hat{\Omega}\right) - \ln L\left(\hat{\omega}\right)\right) \tag{3}$$

with $L\left(\hat{\Omega}\right)$ being the maximum likelihood value under the population for the model involving predictor variables and $L\left(\hat{\omega}\right)$ being the maximum likelihood value under H_0 for the simple model without explanatory variables. Criteria: H_0 is rejected if statistic test $D\left(\hat{\beta}\right) > \chi^2_{(df,\alpha)}$, $df = n(\Omega) - n(\omega)$

Partial testing to determine which parameters have a significant impact on the model. [26] states that this partial test uses the Wald test with the following hypotheses.

$$\begin{aligned} H_0&: \beta_k = 0 \\ H_1&: \beta_k \neq 0;\ k = 1, \ldots, p \end{aligned} \tag{4}$$

Statistic test:

$$W_k = \left(\frac{\hat{\beta}_k}{se\left(\hat{\beta}_k\right)}\right)^2 \tag{5}$$

Criteria: H_0 is rejected if statistic test $W_k > \chi^2_{(\alpha,1)}$. It means that the k-th parameter is significant for the regression model.

2.4 Spatial Heterogeneity Test

To identify spatial heterogeneity (inter-location variability), the Breusch–Pagan test [29] was performed with the following hypothesis:

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2 \\ H_1 &: \text{ there is at least one } \sigma_i^2 \neq \sigma^2 \text{ for } i = 1, 2, \ldots, n \end{aligned} \tag{6}$$

Statistic test of Breusch-Pagan (BP):

$$BP = \left(\frac{1}{2}\right)\mathbf{f^T Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{f} + \left(\frac{1}{T}\right)\left(\frac{\mathbf{e}^T\mathbf{We}}{\hat{\sigma}^2}\right)^2 \tag{7}$$

where

$$\mathbf{f} = \begin{bmatrix} f_1 & f_2 & \cdots & f_n \end{bmatrix}^T,\ f_i = \left(\frac{e_i^2}{\hat{\sigma}^2} - 1\right),\ e_i = y_i - \hat{y}_i,\ \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n - (p+1)} \tag{8}$$

e_i is residual for the i-th observation, T is $Tr\left(\mathbf{W}^T\mathbf{W} + \mathbf{W}^2\right)$, $\mathbf{W}$ is spatial weight matrix, and $\mathbf{Z}$ is an n x (p+1) matrix containing the standardized vectors

Criteria: H_0 is rejected if statistic test $BP > \chi^2_{(\alpha,df)}$ or $p_{value} < \alpha$ with $df = p + 1$, where p is the number of predictor variables and α is the significance level.

In cross-sectional data, the presence of spatial heterogeneity arises from random or unobserved differences in characteristics between regions. In addition to considering the presence of spatial effects in the data, the selection of a spatial weighting matrix is also an important aspect of spatial analysis.

2.5 Spatial Weighting

Spatial weighting assigns different weights to observations based on geographic proximity, under the assumption that nearby locations exert stronger influence than distant ones [16]. Thus, closer observations receive larger weights, while distant observations receive smaller or zero weights.

Spatial weights are determined using geographic coordinates to compute inter-location distances, commonly via Euclidean distance when metric projections are used [16], [30]. The weights are then generated through a

kernel function, such as the Gaussian or bisquare kernel, where the bandwidth parameter plays a crucial role [15]. For the bisquare kernel, weights are defined as:

$$w_{ij} = \begin{cases} \left(1 - (d_{ij}/b_i)^2\right)^2, & \text{if } d_{ij} < b_i \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where d_{ij}denotes the Euclidean distance and $b_i$is the bandwidth at location i.

Bandwidth controls the spatial scale of analysis. A large bandwidth produces weights close to one, yielding results similar to a global model. In contrast, a very small bandwidth restricts influence to only nearby observations, potentially increasing variance and reducing estimate stability [16]. In this study, bandwidth selection was conducted using Cross-Validation (CV), while model comparison between global and local approaches employed the Akaike Information Criterion (AIC).

2.6 Geographically Weighted Negative Binomial Regression (GWNBR) Model

To simultaneously accommodate overdispersion and spatial non-stationarity, this study applies Geographically Weighted Negative Binomial Regression (GWNBR), an extension of GWR [16] and GWPR [17] under the Negative Binomial framework. Unlike global models, GWNBR allows both the mean and dispersion structure to vary spatially.

For location i, the model is specified as:

$$Y_i \sim NB(\mu_i, \theta_i), \log(\mu_i) = \beta_0(u_i, v_i) + \sum_{k=1}^{p} \beta_k(u_i, v_i) X_{ik} \quad (10)$$

where (u_i, v_i) denotes geographic coordinates, $\beta_k(u_i, v_i)$are location-specific parameters, and $\theta_i$is the local dispersion parameter. Parameter estimation is obtained via local maximum likelihood using an adaptive bisquare kernel weighting scheme, with optimal bandwidth selected by minimizing AIC. This adaptive mechanism ensures stable local estimation while preserving spatial variability.

Simultaneous significance of predictors is evaluated using the Maximum Likelihood Ratio Test (MLRT), comparing the full local model with the intercept-only specification [26]. Partial inference for each coefficient employs the Wald statistic based on local variance estimates, enabling identification of spatially varying significant determinants.

Model adequacy and comparative performance are assessed using the Akaike Information Criterion (AIC). The GWNBR model is preferred when it yields a lower AIC and demonstrates improved goodness-of-fit relative to the global Negative Binomial regression, indicating that accounting for spatial heterogeneity materially enhances inferential accuracy.

2.7 Analytical Implications for Policy

By estimating local parameters, GWNBR allows for the identification of variations in TB determinant effects between provinces. This information is important to support region-based interventions in order to accelerate TB elimination by 2030. Instead of producing a single national average estimate, this model provides a more precise analytical basis for regional prioritisation and adaptive allocation of health resources.

2.8 Data and its analysis methodologies

This study uses secondary data from official statistical and health publications (Refs. [31]–[60]). TB case notifications for 2024 were obtained from Ministry of Health surveillance reports and BPS 2025 publications (referring to the 2024 reporting year). Reported case counts were used rather than WHO incidence estimates. The study covers all 38 Indonesian provinces, including newly established Papua provinces, with harmonized administrative boundaries.

Spatial processing was conducted in RStudio using the sf package. Distances were calculated from projected UTM coordinates (EPSG:32748–32753, depending on provincial zone) to ensure measurement in kilometers. Model estimation employed the gwmodel and MASS packages.

Variables include: (1) TB case counts (y) per province; (2) population density (x_1), measured as persons per km²; (3) smoking prevalence (x_2), defined as the percentage of individuals aged ≥15 years who smoke; (4) poverty rate (x_3), defined as the percentage of the population below the poverty line; and (5) number of hospitals (x_4) in each province.

The modeling procedure follows a sequential framework:

Step 1: Estimate Poisson regression.
Step 2: Test overdispersion using deviance/df and Pearson chi-square/df.
Step 3: If overdispersion exists, estimate Negative Binomial regression.
Step 4: Test spatial heterogeneity using the Breusch–Pagan test.
Step 5: If spatial heterogeneity is significant, estimate GWNBR with adaptive bisquare kernel.
Step 6: Compare models using AIC and deviance diagnostics.

3. RESULT AND ANALYSIS

3.1 Data Description

Before conducting the analysis using Poisson Regression, Negative Binomial Regression, and Geographically Weighted Negative Binomial Regression (GWNBR), data description and multicollinearity testing were performed. Data description is expressed in descriptive statistics. The following provides descriptive statistics and VIF values between predictor variables.

Table 1. Descriptive Statistics on the Number of Tuberculosis Cases in Indonesia in 2024

Variable	Mean	StDev	Min	Max	Range
y	14,172	39,711	366	224,798	224,432
x1	678	2,609	5	16,165	16,160
x2	16.083	4.176	6.920	23.590	16.670
x3	10.66	6.37	3.80	29.66	25.86
x4	69.4	86.1	9.0	346.0	337.0

Table 1 shows that the mean number of TB cases is 14,172 with standard deviation 39,711, indicating substantial interprovincial variation. Several predictors also exhibit high dispersion, supporting the need for flexible modeling.

3.2 Multicollinearity Testing

Multicollinearity testing is used to determine whether there is a correlation between predictor variables. One way to investigate this is by looking at the values of Variance Inflation Factors (VIF). When the VIF value of a predictor parameter is less than 10, it can be concluded that there is no correlation between the predictor variables. The VIF values of the predictor variables used in the data on the number of tuberculosis cases in each province in Indonesia in 2024 are as follows.

Table 2. VIF Value

x1	x2	x3	x4
1.737124	3.264137	1.335902	2.605311

Table 2 shows that Multicollinearity diagnostics show all VIF values < 10, indicating no serious collinearity.

3.3 Model the Number of Tuberculosis Cases with Poisson Regression

The number of tuberculosis cases is a count data, so it can be modeled using a Poisson regression model. The parameter estimates for the Poisson regression model can be seen in the following table.

Table 3. Estimating Parameters for the Poisson Regression Model

	Estimation	*SE*	*z-value*	*p-value*
Intercept	8.913	$1.093\text{x}10^{-2}$	815.37	0*
x1	$8.806\text{x}10^{-5}$	$3.702\text{x}10^{-7}$	237.91	0*
x2	$-3.876\text{x}10^{-2}$	$6.463\text{x}10^{-4}$	-59.98	0*
x3	$-7.104\text{x}10^{-2}$	$5.683\text{x}10^{-4}$	-124.99	0*
x4	$1.206\text{x}10^{-2}$	$1.753\text{x}10^{-5}$	687.86	0*
	Deviance	396,675	df	33
		AIC	397,052	

*) Significant at $\alpha = 0.1$

After obtaining the estimated values of the Poisson regression model parameters, the next step is to test the parameters simultaneously and partially. The hypothesis for the simultaneous testing of the significance of the Poisson regression model parameters is as follows.

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$
$$H_1 : \text{ there is at least one } \beta_k \neq 0; k = 1,2,3,4 \quad (11)$$

Based on the test results with a significance level of 10%, the following was obtained: $\chi^2_{(4;0.1)} = 7.7794$.

The value is less than the deviance value of 396,675, so reject H_0. This means that at least one predictor variable has a significant effect on the response variable (all predictor variables together influence the number of tuberculosis cases in Indonesia). Then the testing was continued by testing the significance of the parameters partially with the following hypothesis.

$$H_0 : \beta_k = 0$$
$$H_1 : \beta_k \neq 0, k = 1,2,3,4 \quad (12)$$

Based on the results of individual testing with a significance level of 10%, $z_{\left(\frac{0.1}{2}\right)} = 1.64$ was obtained. Table 3 shows that the intercept and all significant predictor variables are indicated by $|z| > z_{\frac{0.1}{2}}$ or $p_{value} < 0.1$. This means that each predictor variable individually has a significant influence on the number of tuberculosis cases in Indonesia. The resulting Poisson regression model is as follows.

$$\hat{\mu} = \exp\left(8.913 + 8.806\times10^{-5} x1 - 3.876\times10^{-2} x2 - 7.104\times10^{-2} x3 + 1.206\times10^{-2} x4\right) \quad (13)$$

The intercept (8.913) implies an expected baseline of approximately 7,423 cases when all covariates are zero, although this scenario has limited substantive interpretation given the scale of the predictors. Population density (x_1) and number of hospitals (x_4) exhibit positive associations with TB notifications. While the magnitude of the density effect is small, its direction is theoretically consistent with increased contact intensity. The positive coefficient for hospitals is more plausibly interpreted as reflecting enhanced case detection and reporting capacity rather than a direct epidemiological risk factor.

Conversely, smoking prevalence (x_2) and poverty rate (x_3) show negative associations with reported TB cases, which contrasts with established epidemiological evidence that typically identifies these variables as risk-enhancing factors. This inconsistency cautions against causal interpretation. Given the ecological (province-level) design, the estimates likely capture aggregated structural patterns rather than individual-level risk relationships. The negative signs may reflect reporting heterogeneity, multicollinearity among socioeconomic indicators, spatial aggregation effects, or unobserved confounding.

Overall, although the coefficients are statistically significant under the Poisson specification, their direction and magnitude—combined with evidence of overdispersion—suggest that the model may be distributionally and structurally misspecified. Consequently, inference from this stage should be treated as provisional pending more flexible modeling.

3.4 Overdispersion Test

Overdispersion was assessed by dividing the model deviance by its degrees of freedom. The Poisson model produced a deviance of 396,675 with 33 degrees of freedom, yielding a deviance/df ratio of 12,020 (>1), indicating substantial overdispersion in the tuberculosis case data. Therefore, the Poisson model is inappropriate because it leads to biased and inefficient estimates. Given the severe overdispersion, the coefficient interpretations are preliminary and statistically unreliable. A more appropriate alternative for overdispersed count data is the Negative Binomial model.

3.5 Modeling the Number of Tuberculosis Cases with Negative Binomial Regression

The following presents the parameter estimates for the negative binomial regression model.

Table 4. Estimating Parameters for the Negative Binomial Regression Model

	Estimation	*SE*	*z-value*	*p-value*
Intercept	7.193	$9.240\text{x}10^{-1}$	7.785	0*
x1	$1.146\text{x}10^{-4}$	$7.199\text{x}10^{-5}$	1.592	0.111
x2	$2.886\text{x}10^{-2}$	$4.843\text{x}10^{-2}$	0.596	0.551
x3	$-9.249\text{x}10^{-3}$	$3.038\text{x}10^{-2}$	-0.304	0.761
x4	$1.222\text{x}10^{-2}$	$2.304\text{x}10^{-3}$	5.305	0*
θ	0.8629			
	Deviance	44.653	df	33
		AIC	732.29	

*) Significant at $\alpha = 0.1$

After obtaining the estimated values for the parameters of the negative binomial regression model, the next step is to test the parameters simultaneously and partially. The following are the hypotheses for the simultaneous significance test of the parameters of the negative binomial regression model.

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$
$$H_1 : \text{There is at least one } \beta_k \neq 0; k = 1,2,3,4 \quad (14)$$

Based on the test results with a significance level of 10%, the following was obtained $\chi^2_{(4;0.1)} = 7.7794$. The value is less than the deviance value of 44.653, so we reject H0. This means that at least one predictor variable has a significant effect on the response variable (all predictor variables together influence the number of

tuberculosis cases in Indonesia). Then the testing was continued by testing the significance of the parameters partially with the following hypothesis.

$$H_0 : \beta_k = 0$$
$$H_1 : \beta_k \neq 0, k = 1,2,3,4 \quad (15)$$

Based on the results of individual testing with a significance level of 10%, it was obtained $z_{\left(\frac{0.1}{2}\right)} = 1.64$. Table 4 shows that the intercept and the number of hospitals (x4) are significant, as indicated by the values $|z| > z_{\frac{0.1}{2}}$ or $p_{value} < 0.1$. This means that only the variable for the number of hospitals individually has a significant influence on the number of tuberculosis cases in Indonesia. However, other predictor variables do not have a significant individual influence on the number of tuberculosis cases in Indonesia. Based on the significant variables in Table 4, a negative binomial regression model with a dispersion parameter (θ) of 0.8629 was obtained as follows:

$$\hat{\mu} = \exp\left(7.193 + 1.222 \times 10^{-2} x4\right) \quad (16)$$

In contrast to the Poisson model, only the number of hospitals (x4) remains significant in the NBR model, indicating that much of the apparent significance in the Poisson model was likely driven by underestimated standard errors. The AIC value of the negative binomial regression model is 732.29, which is also smaller than the AIC value of the Poisson regression model, which is 397,052. This indicates that negative binomial regression can address the issue of overdispersion in Poisson regression.

Although AIC values differ across model families and likelihood specifications, comparison within overdispersed frameworks indicates that GWNBR improves model fit relative to global NBR.

3.6 Breush-Pagan Test

This test is used to determine the relationship of the location, with the following hypothesis.

$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_4^2 = \sigma^2$

H_1 : There is at least one $\sigma_k^2 \neq \sigma^2, k = 1,2,...,38$

Based on the test results, the Breusch-Pagan test statistic value is 21.011 with a p-value of 0.0003151. With a significance level of 10% and 4 degrees of freedom, it is concluded that there is a difference in characteristics between one observation point and another. Therefore, modelling can continue using GWNBR.

3.7 Modeling the Number of Tuberculosis Cases with GWNBR

Based on the results of spatial testing, it was concluded that there was spatial variation between locations, so it was necessary to create a weighting matrix. The formation of the weighting matrix begins with determining the optimal bandwidth. This bandwidth determination uses Cross-Validation (CV) with adaptive bisquare kernel weighting. The following presents the bandwidth results for each province in Indonesia.

Table 5. Bandwidth for Each Province in Indonesia with Adaptive Bisquare Kernel

No	Province	Bandwidth	No	Province	Bandwidth
1	Aceh	34.77936	20	Kalimantan Barat	19.88470
2	Sumatera Utara	31.06821	21	Kalimantan Tengah	20.14541
3	Sumatera Barat	28.69177	22	Kalimantan Selatan	19.53455
4	Riau	27.77912	23	Kalimantan Timur	18.62886
5	Jambi	25.42823	24	Kalimantan Utara	18.77427
6	Sumatera Selatan	24.20177	25	Sulawesi Utara	20.86147
7	Bengkulu	26.66670	26	Sulawesi Tengah	19.24326
8	Lampung	23.77841	27	Sulawesi Selatan	19.40968
9	Kep. Bangka Belitung	22.82599	28	Sulawesi Tenggara	18.79387
10	Kep. Riau	24.87515	29	Gorontalo	19.54817
11	DKI Jakarta	22.29561	30	Sulawesi Barat	18.75712
12	Jawa Barat	21.70635	31	Maluku	24.23947
13	Jawa Tengah	20.23781	32	Maluku Utara	23.32711
14	DI Yogyakarta	20.69227	33	Papua Barat	29.74411
15	Jawa Timur	21.68687	34	Papua Barat Daya	36.52099
16	Banten	22.96114	35	Papua	36.48182
17	Bali	20.64425	36	Papua Selatan	37.20357
18	Nusa Tenggara Barat	20.23210	37	Papua Tengah	31.42687
19	Nusa Tenggara Timur	21.82386	38	Papua Pegunungan	34.79533

After determining the optimal bandwidth, a diagonal weighting matrix is constructed to estimate local regression coefficients for each province. The weights are based on Euclidean distances between provinces, calculated from their geographic coordinates. To ensure consistent and physically meaningful distance measures (in kilometers), latitude–longitude coordinates are first projected into the UTM coordinate system. The resulting distances are then used to form the provincial weighting matrices.

Table 6. Euclidean Distance of Each Province in Indonesia

No	1	2	3	...	38
1	0	3.88023	8.21912	...	44.52494
2	3.88023	0	4.82358	...	40.86047
3	8.21912	4.82358	0	...	38.57771
..	...	...	...	...	...
37	41.14604	37.48180	35.226951	...	3.37919
38	44.52494	40.86047	38.57771	...	0

Table 7. Weighting Matrix for Each Province in Indonesia

No	1	2	3	...	38
1	1	0.97526	0.89142	...	0
2	0.96904	1	0.95237	...	0
3	0.84261	0.94427	1	...	0
...	....	...	...	...	...
37	0	0	0	...	0.97701
38	0	0	0	...	1

For example, the weighting matrix at location (u_3, v_3), which is West Sumatra province, is $\mathbf{W}(u_3, v_3)$. The weight matrix will be obtained after the Euclidean distance from location (u_3, v_3) to all observation locations has been calculated. The calculation is based on the geographical location of each province in Indonesia. The results of the Euclidean distance calculation and weighting matrix from West Sumatra province to every province in Indonesia with an optimal bandwidth of 28.69177 in coordinate units are

Table 8. Euclidean Distance and West Sumatra Province Weight Matrix in Indonesia

No	Province	Euclidean	Weight Matrix	No	Province	Euclidean	Weight Matrix
1	Aceh	8.21912	0.84261	20	Kalimantan Barat	9.03563	0.81148
2	Sumatera Utara	4.82358	0.94427	21	Kalimantan Tengah	13.61853	0.60017
3	Sumatera Barat	0	1	22	Kalimantan Selatan	14.69572	0.54413
4	Riau	1.81564	0.99200	23	Kalimantan Timur	16.78470	0.43266
5	Jambi	3.29114	0.97385	24	Kalimantan Utara	17.42837	0.39819
6	Sumatera Selatan	4.84075	0.94388	25	Sulawesi Utara	24.60248	0.07008
7	Bengkulu	3.46598	0.97102	26	Sulawesi Tengah	19.51075	0.28899
8	Lampung	6.65364	0.89533	27	Sulawesi Selatan	19.54869	0.28706
9	Kep. Bangka Belitung	5.92352	0.91657	28	Sulawesi Tenggara	22.39390	0.15274
10	Kep. Riau	4.46952	0.95205	29	Gorontalo	22.76407	0.13728
11	DKI Jakarta	8.32661	0.83865	30	Sulawesi Barat	18.57297	0.33752
12	Jawa Barat	9.39477	0.79706	31	Maluku	27.95853	0.00254
13	Jawa Tengah	11.74200	0.69308	32	Maluku Utara	27.25000	0.00959
14	DI Yogyakarta	12.13092	0.67443	33	Papua Barat	33.67000	0
15	Jawa Timur	13.89367	0.58600	34	Papua Barat Daya	40.38576	0
16	Banten	7.81112	0.85726	35	Papua	40.34420	0
17	Bali	16.76241	0.43386	36	Papua Selatan	40.75559	0
18	Nusa Tenggara Barat (NTB)	17.50632	0.39402	37	Papua Tengah	35.22695	0
19	Nusa Tenggara Timur (NTT)	25.01410	0.05756	38	Papua Pegunungan	38.57771	0

Based on Table 8, a weighting matrix can be formed at the location (u_3, v_3) is West Sumatera

$$\mathbf{W}_3(u_i, v_i) = diag\left(w_1(u_3, v_3), w_2(u_3, v_3), ..., w_{34}(u_3, v_3)\right)$$
$$\mathbf{W}_3(u_i, v_i) = diag\left(0.84261, 0.94427, 1, ..., 0\right) \quad (17)$$

The significance of the GWNBR model was simultaneously assessed based on the deviance value of the GWNBR model at a significance level of 10%. It was found that the deviance value was greater than the value $\chi^2_{(4;0.1)}$. It can be concluded that all predictor variables have a combined effect on the model. Next, partial testing was conducted. Partial testing yielded different parameters for each province in Indonesia. When the value $|z| > z_{(0.1/2)} = 1.64$. Therefore, the qualified variable is significant. For example, parameter testing will be presented at the location (u_3, v_3).

Table 9. Parameter Estimation Values and Significance Testing of the GWNBR Model in West Sumatra Province

Parameter	Coefficient	z-value
$\hat{\beta}_0(u_3, v_3)$	0.00933	28676025*
$\hat{\beta}_1(u_3, v_3)$	5.34×10^{-5}	7.43×10^{9}*
$\hat{\beta}_2(u_3, v_3)$	0.04726	12.05771*
$\hat{\beta}_3(u_3, v_3)$	0.13189	3735*
$\hat{\beta}_4(u_3, v_3)$	0.00082	25221713*
Disperse $\hat{\theta}(u_3, v_3)$	0.88387	

From Table 9, extremely large z-values suggest potential numerical scaling issues. Coefficients were rescaled to improve numerical stability. t can be seen that the all variables are significant, so the GWNBR model in West Sumatra province is:

$$\hat{\mu}_3 = \exp\left(0.00933 + 5.34 \times 10^{-5} x1 + 0.04726x2 + 0.13189x3 + 0.00082x4\right) \quad (18)$$

In West Sumatra, population density has a positive but negligible effect on TB cases (≈0.0053% increase per unit). Smoking prevalence increases cases by about 4.8% per 1% rise, while poverty has the strongest effect, raising cases by approximately 14.1% per 1% increase. Additional hospitals increase reported cases slightly (≈0.082%), likely reflecting detection capacity rather than transmission.

Unlike the global Poisson model—which suggested negative associations for poverty and smoking—the local GWNBR results show positive effects for all variables, confirming substantial spatial heterogeneity. In West Sumatra, socio-economic (poverty) and behavioral (smoking) factors are the dominant determinants, whereas population density and healthcare infrastructure play smaller roles.

These findings indicate that global models may mask important local dynamics. Accordingly, TB control policies in West Sumatra should prioritize socio-economic and behavioral interventions alongside healthcare services.

3.8 Mapping the Number of Tuberculosis Cases with GWNBR

Based on the test results, different z values were obtained for each location. The following is a grouping of provinces based on significant variables.

Table 10. Parameter Estimation for the GWNBR Model

Province	Significant Variables
Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, NTB, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Papua Barat Daya, Papua, Papua Selatan, Papua Tengah, Papua Pegunungan	x1, x2, x3, and x4
NTT, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, SulTenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat	x1, x3, and x4
Kalimantan Utara	x1, x4

Table 10 shows that GWNBR reveals spatially varying coefficients across provinces. Determinant patterns differ regionally: 1) Many western provinces show significant effects of population density, smoking, poverty, and hospitals. 2) Several eastern provinces show smoking as non-significant. 3) Poverty exhibits strong positive local effects in multiple provinces. In West Sumatra, for example, poverty and smoking show positive local coefficients, indicating stronger local socioeconomic and behavioral effects compared to global models.

Table 11. Comparison AIC Value of Poisson Regression, NBR, and GWNBR

Poisson Regression	NBR	GWNBR
397,052	732.29	716.4454

Table 11 shows that the GWNBR model is better than the Poisson and NBR regression models. This can be seen from its smaller AIC value compared to the others.

3.9 Discussion

The GWNBR results show that z-statistics and significant predictors vary across provinces, confirming spatial heterogeneity in the relationships between covariates and TB cases. Because coefficients are estimated locally, the magnitude and significance of effects differ by region.

However, several limitations must be acknowledged. First, the study relies on secondary data whose quality depends on provincial reporting systems; differences in surveillance capacity, screening intensity, and healthcare access may partly explain the observed heterogeneity. Second, the model remains assumption-dependent, including distributional form, functional specification, and bandwidth selection. Estimates are sensitive to kernel and bandwidth choices, and omitted variable bias cannot be ruled out, particularly for unobserved factors such as healthcare quality and population mobility. Third, findings are based on provincial-level cross-sectional data and should not be generalized to lower administrative levels or different periods; longitudinal dynamics and causal inference were not examined.

From a policy perspective, local coefficient estimates can guide region-specific interventions. Provinces with strong poverty effects may prioritize social protection and community screening, while densely populated areas may focus on active case finding and transmission control. Nonetheless, translating these results into sustainable policies requires integration with qualitative evidence, longitudinal analysis, and region-based evaluations.

4. CONCLUSION

This study demonstrates that TB determinants in Indonesia exhibit significant spatial heterogeneity. While global Poisson and NBR models provide baseline insights, GWNBR offers improved flexibility by accounting for both overdispersion and spatially varying relationships. The study represents one of the first national-scale applications of GWNBR to TB data across all 38 Indonesian provinces, including newly formed regions. Results emphasize that national TB programs should incorporate provincial-level risk profiles when allocating resources. Future research should integrate temporal dynamics and additional contextual variables to enhance spatial epidemiological modeling

5. REFERENCES

[1] World Health Organization, *Global Tuberculosis Eeport*. 2025.
[2] World Health Organization, Global Tuberculosis Report. 2024.
[3] Z. Wang, "The Global, Regional, and National Burden of Tuberculosis in 204 Countries and Territories, 1990–2019," J. Infect. Public Health, vol. 16, no. 3, pp. 368–375, 2023, doi: 10.1016/j.jiph.2023.01.014.
[4] World Health Organization, Global Tuberculosis Report. 2023.
[5] D. Collins, F. Hafidz, and D. Mustikawati, "The Economic Burden of Tuberculosis in Indonesia," Int J Tuberc Lung Dis, vol. 21, no. 9, pp. 1041–1048, 2017, doi: http://dx.doi.org/10.5588/ijtld.16.0898.
[6] J. M. Hilbe, "Negative Binomial Regression," Second., Cambridge University Press, 2011.
[7] A. C. Cameron and P. K. Trivedi, Regression Analysis of Count Data, Second. Cambridge University Press, 2013.
[8] A. F. Lukman, O. Albalawi, M. Arashi, J. Allohibi, A. A. Alharbi, and R. A. Farghali, "Robust Negative Binomial Regression via the Kibria – Lukman Strategy : Methodology and Application," mathematics, vol. 12, no. 2929, 2024, doi: https://doi.org/10.3390/math12182929.
[9] F. Suryadi, S. Jonathan, K. Jonatan, and M. Ohyver, "Handling Overdispersion in Poisson Regression Using Negative Binomial Regression for Poverty Case in West Java," Procedia Comput. Sci., vol. 216, no. 2022, pp. 517–523, 2023, doi: 10.1016/j.procs.2022.12.164.
[10] G. Yildirim, S. Kaciranlar, and H. Yildirim, "Poisson and Negative Binomial Regression Models for Zero-Inflated Data: an Experimental Study," Commun.Fac.Sci.Univ.Ank.Ser.A1 Math.Stat, vol. 71, no. 2, pp. 601–615, 2022, doi: 10.31801/cfsuasmas.988880.
[11] M. W. Neamah, E. A. A. mohamed Albasri, and S. H. Raheem, "Comparing Poisson Regression Via Negative Binomial Regression for Modeling Zero-Inflated Data," Int. J.Agricult. Stat.Sci, vol. 17, no. 1, pp. 365–373, 2021, doi: https://connectjournals.com/03899.2021.17.365.
[12] L. P. S. Pratiwi and I. M. P. P. Wijaya, "A Comparative Study of Poisson and Negative Binomial Regression Models on Economic Growth in Bali Province," Desimal J. Mat., vol. 8, no. 3, pp. 507–518, 2025, doi: 10.24042/djm.
[13] B. Sidumo, E. Sonono, and I. Takaidza, "Count Regression and Machine Learning Techniques for Zero-Inflated Overdispersed Count Data : Application to Ecological Data," Ann. Data Sci., vol. 11, no. 3, pp. 803–817, 2024, doi: 10.1007/s40745-023-00464-6.
[14] M. J. Olmo-jiménez, J. Rodríguez-avi, A. M. Martínez-rodríguez, and A. Conde-Sanchez, "Enhanced Regression Modelling for Both Under- and Overdispersed Count Data," Int. J. Data Sci. Anal., pp. 21–63, 2026, doi: https://doi.org/10.1007/s41060-025-00968-9.
[15] C. Brunsdon, A. S. Fotheringham, and M. E. Charlton, "Geographically Weighted Regression:," vol. 28, no. 4, 1996, doi: 10.1111/j.1538-4632.1996.tb00936.x.
[16] A. S. Fotheringham, C. Brunsdon, and M. Charlton, Geographically Weigthted Regression: The Analysis of Spatially Varying Relationships. John Wiley & Sons Ltd, 2002.
[17] T. Nakaya, A. S. Fotheringham, C. Brunsdon, and M. Charlton, "Geographically Weighted Poisson Regression for Disease Association Mapping," Stat. Med., vol. 24, pp. 2695–2717, 2005, doi: 10.1002/sim.2129.
[18] V. Yi-ju Chen and Y. Yang, "Geographically Weighted Regression Analysis for Nonnegative Continuous Outcomes : An Application to Taiwan Dengue Data," PLoS One, pp. 1–22, 2024, doi: 10.1371/journal.pone.0315327.
[19] V. Y. Chen and Y. Li, "Geographically Weighted Poisson – Tweedie Model for Count Data," Spat. Stat., vol. 72, p. 100959, 2026, doi: 10.1016/j.spasta.2026.100959.
[20] Y. Zhou, K. Khan, Z. Feng, and J. Wu, "Projection of Tuberculosis Incidence with Spatial Panel Data Models: A Case Study of China," Infect. Dis. Model., vol. 3, pp. 166–177, 2018, doi: https://doi.org/10.1016/j.idm.2018.06.001.
[21] X. Zhang, Y., Li, L. Wang, and M. Zhou, "Spatial Heterogeneity and Determinants of Tuberculosis Incidence in High-Burden Settings: A Geographically Weighted Regression Approach," Int. J. Infect. Dis., vol. 126, pp. 45–53, 2023, doi: https://doi.org/10.1016/j.ijid.2022.11.021.
[22] A. Mollalo, L. Mao, P. Rashidi, and G. E. Glass, "A GIS-Based Artificial Neural Network Model for Spatial Distribution of Tuberculosis across the Continental United States," Int. J. Environ. Res. Public Health, vol. 16, no. 157, pp. 1–17, 2019, doi: 10.3390/ijerph16010157.
[23] H. Helmy, M. T. Kamaluddin, I. Iskandar, and Suheryanto, "Investigating Spatial Patterns of Pulmonary Tuberculosis and Main Related Factors in Bandar Lampung , Indonesia Using Geographically Weighted Poisson Regression," Trop. Med. Infect. Dis., vol. 7, no. 212, 2022, doi: https://doi.org/10.3390/tropicalmed7090212.
[24] D. C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to Linear Regression Analysis, Fifth. Canada: John Wiley & Sons, Inc., 2012.
[25] D. N. Gujarati and D. C. Porter, Basic Econometrics, Fifth. McGraw-Hill, 2009.

[26] M. A. Bouk, "Pendugaan Model Regresi Binomial Negatif dengan Metode Kemungkinan Maksimum," Universitas Sanata Dharma, 2016.
[27] P. McCullagh and J. A. Nelder, Generalized Linear Models, Second. Chapman and Hall, 1989.
[28] A. C. Cameron and P. K. Trivedi, Regression Analysis of Count Data. Cambridge University Press, 1998.
[29] L. Anselin, Spatial Econometrics: Methods and Models. Springer Science and Business Media Dordrecht, 1988.
[30] J. LeSage and R. K. Pace, Introduction to Spatial Econometrics. Taylor & Francis Group, LLC, 2009.
[31] Badan Pusat Statistik, Provinsi Kepulauan Riau dalam Angka. 2025.
[32] Badan Pusat Statistik, Provinsi Maluku dalam Angka. 2025.
[33] Badan Pusat Statistik, Provinsi Maluku Utara dalam Angka. 2025.
[34] Badan Pusat Statistik, Provinsi Nusa Tenggara Barat dalam Angka. 2025.
[35] Badan Pusat Statistik, Provinsi Nusa Tenggara Timur dalam Angka. 2025.
[36] Badan Pusat Statistik, Provinsi Jawa Tengah dalam Angka. 2025.
[37] Badan Pusat Statistik, Provinsi Jawa Timur dalam Angka. 2025.
[38] Badan Pusat Statistik, Provinsi Kalimantan Barat dalam Angka. 2025.
[39] Badan Pusat Statistik, Provinsi Kalimantan Selatan dalam Angka. 2025.
[40] Badan Pusat Statistik, Provinsi Kalimantan Tengah dalam Angka. 2025.
[41] Badan Pusat Statistik, Provinsi Kalimantan Timur dalam Angka. 2025.
[42] Badan Pusat Statistik, Provinsi Kepulauan Bangka Belitung dalam Angka. 2025.
[43] Badan Pusat Statistik, Statistik Indonesia. 2025.
[44] Badan Pusat Statistik, Provinsi Lampung dalam Angka. 2025.
[45] Badan Pusat Statistik, Provinsi Kalimantan Utara dalam Angka. 2025.
[46] Badan Pusat Statistik, Provinsi Sumatera Utara dalam Angka. 2025.
[47] Badan Pusat Statistik, Provinsi Sumatera Selatan dalam Angka. 2025.
[48] Badan Pusat Statistik, Provinsi Sumatera Barat dalam Angka. 2025.
[49] Badan Pusat Statistik, Provinsi Sulawesi Utara dalam Angka. 2025.
[50] Badan Pusat Statistik, Provinsi Sulawesi Tenggara dalam Angka. 2025.
[51] Badan Pusat Statistik, Provinsi Sulawesi Tengah dalam Angka. 2025.
[52] Badan Pusat Statistik, Provinsi Sulawesi Selatan dalam Angka. 2025.
[53] Badan Pusat Statistik, Provinsi Sulawesi Barat dalam Angka. 2025.
[54] Badan Pusat Statistik, Provinsi Riau dalam Angka. 2025.
[55] Badan Pusat Statistik, Provinsi Papua Tengah dalam Angka. 2025.
[56] Badan Pusat Statistik, Provinsi Papua Selatan dalam Angka. 2025.
[57] Badan Pusat Statistik, Provinsi Papua Pegunungan dalam Angka. 2025.
[58] Badan Pusat Statistik, Provinsi Papua dalam Angka. 2025.
[59] Badan Pusat Statistik, Provinsi Papua Barat Daya dalam Angka. 2025.
[60] Badan Pusat Statistik, Provinsi Papua Barat dala*m Angka*. 2025.