Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 9, Number 3, 2025
DOI: 10.30829/zero.v9i3.26784
Page: 794-806

Determinants of Dengue Hemorrhagic Fever in Aceh: A Panel Regression Approach

[1] Fitra Muliani
Department of Mathematics Education, Universitas Samudra, Kota Langsa, 24416, Indonesia

[2] Fazrina Saumi
Department of Mathematics Education, Universitas Samudra, Kota Langsa, 24416, Indonesia

[3] Amelia
Department of Mathematics Education, Universitas Samudra, Kota Langsa, 24416, Indonesia

[4] Rizki Amalia
Department of Mathematics Education, Universitas Samudra, Kota Langsa, 24416, Indonesia

Article Info

Article history:

Accepted, 26 December 2025

Keywords:

Dengue Hemorrhagic Fever;
Fixed Effect Model;
Panel Data Regression;
Robust Estimation.

ABSTRACT

Dengue Hemorrhagic Fever (DHF) exhibits substantial variation across districts and over time in Aceh Province, making it suitable for analysis within a panel data framework. This study models district-level DHF incidence using applied econometric techniques based on non-spatial panel data regression, employing a balanced panel dataset of 23 districts/cities observed from 2020 to 2022. The Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM) are estimated and formally compared using the Chow test, Hausman test, and Lagrange Multiplier test, with results consistently indicating that the Fixed Effect Model is the most appropriate specification due to the presence of unobserved, time-invariant district-specific effects. Diagnostic testing identifies heteroskedasticity in the error structure; therefore, the selected FEM is re-estimated using White cross-section robust standard errors to ensure reliable statistical inference. Empirical results show that population density is positively and statistically significantly associated with DHF cases, while the number of health workers is negatively and significantly associated, whereas rainfall, number of hospitals, sanitation coverage, and poverty level do not exhibit statistically significant effects in the final robust specification. The selected model explains approximately 86% of the within-district variation in DHF incidence, demonstrating the importance of appropriate model specification and robust variance estimation in panel data regression applied to epidemiological outcomes, while emphasizing that the estimated relationships represent statistical associations rather than causal effects.

Corresponding Author:

Amelia,
Department of Mathematics,
Universitas Samudra, Kota Langsa, Indonesia
Email: amelia@unsam.ac.id

1. INTRODUCTION

Infectious diseases are health conditions caused by microorganisms such as bacteria, viruses, parasites, or fungi, which can be transmitted directly or indirectly between individuals [1]. One of the most prevalent vector-borne infectious diseases in tropical regions is Dengue Hemorrhagic Fever (DHF), which is primarily transmitted by *Aedes* mosquitoes [2]. Dengue continues to pose a major public health challenge in many countries, including Indonesia, due to its wide geographic distribution, recurring outbreaks, and sensitivity to environmental, demographic, and behavioral factors [3].

At the national level, dengue incidence in Indonesia has shown substantial temporal fluctuations, with major outbreaks recorded in several periods, including 2004 and 2016 [6]. Although reported cases declined in subsequent years, dengue remains endemic, with a considerable number of cases still reported by 2020 and beyond [7], [8]. Spatially, dengue cases are unevenly distributed across provinces, with higher concentrations often observed in densely populated and urbanized regions. This heterogeneity suggests that dengue incidence is influenced not only by climatic conditions but also by demographic structure, population mobility, and the distribution of health resources [24], [26].

Aceh Province has experienced a particularly sharp increase in DHF cases in recent years. Official records indicate that reported cases rose from 366 cases in 2021 to more than 2,000 cases in 2022 [9], [10]. This rapid escalation highlights the importance of examining district-level variations within Aceh, as differences in population density, health service availability, and local environmental conditions may contribute to divergent dengue patterns across districts. Understanding these variations is essential for both effective disease control and the development of reliable statistical models that capture regional heterogeneity.

Previous studies have investigated determinants of dengue incidence using various quantitative approaches. Boleng et al. [7] identify population density, poverty, rainfall, and household hygiene practices as important factors, while Winandar and Wati [11] emphasize population density, poverty level, health facilities, health workers, and sanitation access. Other research highlights sociodemographic and environmental characteristics, such as education level, nutritional status, age, vector presence, living environment, and community behaviors including the 3M practice, as contributors to DHF incidence [12], [13]. Several of these studies employ panel regression methods to exploit regional and temporal variation in dengue data.

However, despite the growing use of panel data techniques in dengue research, many existing studies place greater emphasis on epidemiological interpretation than on econometric specification. In particular, earlier panel-based analyses often provide limited justification for estimator choice, insufficient discussion of unobserved regional heterogeneity, and minimal attention to violations of classical assumptions such as heteroskedasticity. As a result, the statistical validity and robustness of inference in some panel dengue studies remain unclear. Moreover, panel analyses focusing specifically on Aceh Province are relatively limited, especially those using recent post-pandemic data [27].

To address these gaps, this study adopts a non-spatial panel data regression framework to analyze district-level DHF cases in Aceh Province. The analysis explicitly compares the Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM), with model selection conducted through formal specification tests, namely the Chow test, Hausman test, and Lagrange Multiplier test. In addition, potential violations of classical assumptions—particularly heteroskedasticity—are explicitly addressed through the application of White cross-section robust standard errors. By using a balanced panel of 23 districts/cities observed over the period 2020–2022, the study captures short-term temporal variation during a critical post-pandemic phase that has not been systematically examined in prior panel-based dengue studies in Aceh.

Accordingly, this study is guided by research questions formulated within a panel data regression framework: (1) which district-level factors are statistically associated with variations in DHF cases over time after controlling for unobserved, time-invariant heterogeneity; (2) whether population density exhibits a positive association with DHF incidence, consistent with increased mosquito–human contact in densely populated areas; and (3) whether the availability of health workers plays a mitigating role, reflected by a negative association with reported DHF cases. From a mathematical and econometric perspective, these questions are addressed through the signs, magnitudes, and statistical significance of the estimated panel regression coefficients under the selected model specification. The analysis is explicitly designed to identify statistical associations rather than causal effects, thereby linking the public health motivation with a rigorously specified panel data modeling objective.

Panel data econometric methods have been widely applied in health economics and disease surveillance to control for unobserved regional heterogeneity and to exploit both cross-sectional and temporal variation in health outcomes [21], [25]. In the context of infectious diseases, such approaches are particularly useful when longitudinal experimental data are unavailable and analysis must rely on routinely collected administrative records [22]. Accordingly, this study adopts a non-spatial panel data regression framework to examine district-level dengue fever patterns in Aceh Province, while explicitly acknowledging the ecological and observational nature of the data.

2. RESEARCH METHOD

This study employs a quantitative research approach using secondary panel data obtained from the Aceh Health Office (Aceh Health Profile) and the Aceh Provincial Statistics Agency (BPS). The dependent variable (Y) is the number of dengue hemorrhagic fever (DHF) cases in each district/city of Aceh Province. The independent variables include: (1) Population density (X1); (2) Rainfall (X2); (3) Number of hospitals/health centers (X3); (4) Percentage of households using proper sanitation (X4); (5) Number of people living in poverty (X5); and (6) Number of health workers (X6).

2.1 Panel Structure, Time Frame, and Variable Construction

This study employs a balanced panel dataset consisting of 23 districts/cities in Aceh Province ($i = 1,2,\dots,23$) observed over three annual periods ($t = 2020,2021,2022$), resulting in a total of $N \times T$observations. Secondary data are obtained from the Aceh Health Office (Aceh Health Profile) and the Aceh Provincial Statistics Agency (BPS). The selected period represents the most recent years for which complete and consistent DHF surveillance and socioeconomic data are available across all districts, while also capturing short-term temporal variation during the post-COVID-19 phase.

The dependent variable, Dengue Hemorrhagic Fever (DHF), is measured as the annual number of reported DHF cases in each district/city and treated as a count variable. No rate transformation (e.g., cases per population) is applied, as population density is explicitly included as a regressor to capture population concentration effects. All explanatory variables are measured annually at the district level and enter the model in their original scale without logarithmic transformation. Variables are not lagged or averaged across years; instead, contemporaneous values are used to maintain consistency with annual surveillance reporting and to avoid imposing additional dynamic assumptions given the short time dimension ($T = 3$).

Population density is measured as the number of individuals per square kilometer. Rainfall is measured in millimeters per year. Health infrastructure is proxied by the number of hospitals or health facilities in each district. Sanitation is measured as the percentage of households with access to adequate sanitation facilities. Poverty is measured as the total number of individuals living below the official poverty line. Health workforce availability is measured as the total number of health workers in each district. All variables are constructed consistently across districts and years based on official definitions provided by BPS and the Aceh Health Office.

To facilitate replication and clarify model interpretation, Table 1 summarizes the definition, unit of measurement, data source, and expected sign of each variable based on prior literature.

Table 1. Definition and Measurement of Variables

Variable	Symbol	Definition	Unit	Data Source	Expected Sign
DHF cases	Y_{it}	Annual reported DHF cases in district i, year t	Number of cases	Aceh Health Office	–
Population density	$X_{1,it}$	Population per unit area	Persons/km²	BPS	+
Rainfall	$X_{2,it}$	Annual average rainfall	mm/year	BPS	±
Health facilities	$X_{3,it}$	Number of hospitals/health centers	Units	BPS	−
Sanitation	$X_{4,it}$	Households with adequate sanitation	Percent (%)	BPS	−
Poverty	$X_{5,it}$	Population below poverty line	Persons	BPS	±
Health workers	$X_{6,it}$	Number of health workers	Persons	BPS	−

2.2 Data Quality, Reporting Limitations, and Pre-processing

DHF surveillance data are subject to several inherent limitations, including potential under-reporting, delayed case notification, and variation in diagnostic and reporting practices across districts and over time. To address these issues and improve data reliability, a series of data quality and pre-processing procedures were implemented prior to model estimation.

First, annual DHF case counts were cross-checked between the Aceh Health Profile and national health statistics to ensure consistency across official sources. Any discrepancies were examined at the aggregate level, and values were retained when differences could be attributed to reporting updates rather than clear errors. Second, district-level time-series plots were inspected to identify abrupt increases or decreases in reported DHF cases. Observations exhibiting extreme values were not automatically classified as outliers or excluded, as such fluctuations may reflect genuine outbreak dynamics. Data points were removed only if there was clear evidence of reporting or transcription errors; no arbitrary statistical thresholds (e.g., trimming or winsorization) were applied. Third, the panel dataset was examined for missing observations. As the final dataset contains complete information for all variables across districts and years, no imputation procedures were required. This avoids the introduction of additional modeling assumptions associated with missing-data correction techniques. Fourth, all variables were harmonized to ensure consistency of operational definitions across data sources and over time, based on official documentation provided by the Aceh Health Office and BPS.

These pre-processing decisions were guided by the objective of preserving the epidemiological integrity of the data while minimizing measurement error. By explicitly stating the criteria for data retention and exclusion, the study avoids subjective data manipulation and enhances transparency and replicability of the panel regression analysis.

2.3 Unit of Analysis and Aggregation Considerations

All variables are aggregated at the district/city level. This introduces potential ecological bias, as district-level associations may not reflect individual-level risk. Furthermore, districts with small populations or low case counts may exhibit unstable annual DHF numbers due to rate fluctuation. These limitations are noted to guide interpretation: the results represent population-level patterns rather than individual-level causal effects.

2.4 Analytical Procedure

Panel data regression was applied using three estimation models: Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM). Model selection was conducted using: Chow test (CEM vs. FEM), Hausman test (FEM vs. REM), and Lagrange Multiplier test (CEM vs. REM).

After model selection, classical assumption tests were performed, including multicollinearity and heteroscedasticity. Significance testing consisted of the F-test, t-test, and interpretation of the adjusted R^2.

The full research flow is shown in Figure 1.

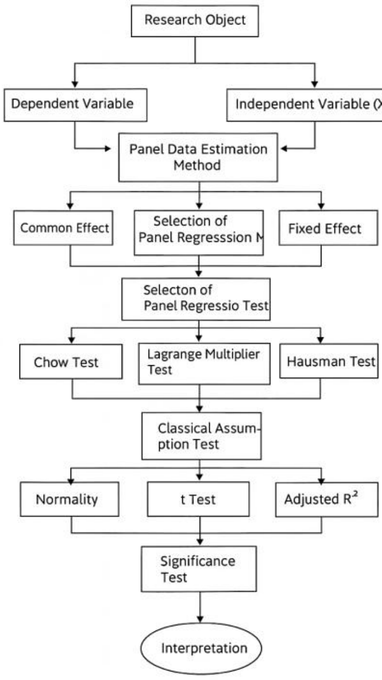

Figure 1. Research flow of the panel data regression method

Under the Fixed Effect Model (FEM), coefficient estimation relies on within-unit variation and remains consistent under the assumption of strict exogeneity between the regressors and the idiosyncratic error term. In panel data settings with a relatively small time dimension and a moderate number of cross-sectional units, the consistency and asymptotic normality of the FEM estimator do not depend on the normality of the error distribution. Moreover, potential serial correlation affects estimator efficiency rather than bias. Consequently, subsequent diagnostic considerations focus on variance robustness rather than distributional assumptions, and inference is conducted using heteroskedasticity-robust standard errors, as detailed in the following section.

3 RESULT AND ANALYSIS

This section presents the empirical results of the panel data regression analysis. The analysis proceeds in several stages, including multicollinearity diagnostics, estimation of alternative panel regression models (CEM, FEM, and REM), model selection tests, diagnostic tests for heteroskedasticity, and interpretation of the final selected model. All analyses are conducted at the district/city level in Aceh Province using balanced panel data.

3.1 Multicollinearity Diagnostics

Prior to estimating the panel regression models, multicollinearity among the explanatory variables was examined using the Variance Inflation Factor (VIF). As shown in Table 2, all centered VIF values are below the commonly accepted threshold of 10, indicating that multicollinearity is not a concern in this study [15]. Therefore, all explanatory variables were retained for subsequent model estimation.

Table 2. Variance Inflation Factors

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
X1	0.019282	40.88825	1.040967
X2	670.2842	3.896297	1.842635
X3	9.651129	21.55919	1.119027
X4	4.683270	3.234961	1.077679
X6	0.002261	3.017555	1.752014
C	156914.7	55.10080	NA

3.2 Panel Regression Model Estimation

Three alternative panel data regression specifications were estimated: the Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM). The dependent variable is the number of dengue hemorrhagic fever (DHF) cases, while the explanatory variables include population density (X_1), rainfall (X_2), number of hospitals/health facilities (X_3), percentage of households with adequate sanitation (X_4), number of people living in poverty (X_5), and number of health workers (X_6).

The estimation results for the CEM, FEM, and REM are reported in Tables 3, 4, and 5, respectively. Coefficient magnitudes and statistical significance vary across specifications, underscoring the importance of formal model selection tests to identify the most appropriate estimator.

a. Common Effect Model (CEM)
The CEM assumes that all cross-sectional units share the same intercept and slope parameters, ignoring unobserved heterogeneity across districts.

Table 3. Estimation Results of the Common Effect Model (CEM)

Variable	Coefficient	Prob.
X1	0.034496	0.0117
X2	-0.039915	0.0066
X3	6.840402	0.0818
X4	-0.891365	0.0067
X5	0.321004	0.1973
X6	-0.009793	0.0644
C	172.5025	0.0001

The results of the CEM model equation based on Equation (1) are as follows.

$$Y_{it} = 172.502 + 0.034X_{1it} - 0.039X_{2it} + 6.840X_{3it} - 0.891X_{4it} - 0.321X_{5it} - 0.009X_{6it} \tag{1}$$

b. Fixed Effect Model (FEM)
The FEM accounts for unobserved time-invariant heterogeneity by allowing each district to have its own intercept. This model is particularly appropriate when district-level characteristics, such as health infrastructure, urban structure, or environmental conditions, may influence DHF patterns but remain constant over time.

Table 4. Estimation Results of the Fixed Effect Model (FEM)

Variable	Coefficient	Prob.
X1	2.400597	0.0001
X2	-0.020738	0.0602
X3	7.831230	0.6111
X4	-1.079370	0.0054
X5	-0.961965	0.5828
X6	-0.025275	0.0000
C	-653.6601	0.0027

The results of the FEM model equation based on Equation (2) are as follows.

$$Y_{it} = -653.66 + 2.401X_{1it} - 0.021X_{2it} + 7.831X_{3it} - 1.079X_{4it} - 0.962X_{5it} - 0.025X_{6it} \quad (2)$$

where α_i represents the district-specific fixed effect.

c. Random Effect Model (REM)
Unlike FEM, the REM assumes that the unobserved individual effects are random and uncorrelated with the explanatory variables.

Table 5. Estimation Results of the Random Effect Model (REM)

Variable	Coefficient	Prob.
X1	0.040553	0.0003
X2	-0.038642	0.0002
X3	6.364961	0.0437
X4	-0.991244	0.0001
X5	0.323854	0.1121
X6	-0.013317	0.0008
C	180.9669	0.0000

The results of the FEM model equation based on Equation (3) are as follows.

$$Y_{it} = 180.967 + 0.041X_{1it} - 0.039X_{2it} + 6.365X_{3it} - 0.991X_{4it} + 0.324X_{5it} - 0.013X_{6it} \quad (3)$$

where μ_i is the random individual effect and ϵ_{it} is the idiosyncratic error.

3.3 Panel Model Selection Tests

To determine the most appropriate panel data model, a sequence of standard specification tests was conducted.

3.3.1 Chow Test (CEM vs. FEM)

The Chow Test examines whether individual fixed effects are needed or whether a pooled model (CEM) is sufficient [16]. The Chow test was applied to examine whether unobserved individual (district-level) effects are jointly significant.
Hypotheses:
H_0: Common Effect Model is appropriate (no fixed effects).
H_1: Fixed Effect Model is appropriate (fixed effects present).
The test results reported in Table 6 show that the p-value of the cross-section F statistic is less than 0.05. Therefore, the null hypothesis is rejected, indicating that the Fixed Effect Model is preferred over the Common Effect Model.

Table 6. Chow Test Result

Effects Test	Statistic	d.f.	Prob.
Cross-section F	4.123735	(22,36)	0.0001
Cross-section Chi-square	81.801076	22	0.0000

3.3.2 Hausman Test (FEM vs. REM)

The Hausman test was then conducted to choose between FEM and REM. The Hausman test evaluates whether the unobserved individual effects are correlated with the regressors.
Hypotheses:
H_0: Random effects are uncorrelated with regressors (REM is appropriate).
H_1: Random effects are correlated with regressors (FEM is appropriate).
As shown in Table 7, the Hausman test yields a p-value below 0.05, leading to rejection of the null hypothesis. This result indicates that the Fixed Effect Model is more appropriate than the Random Effect Model.

Table 7. Correlated Random Effects - Hausman Test

Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Cross-section random	57.067203	6	0.0000

3.3.3 Lagrange Multiplier (LM) Test (CEM vs. REM)

Finally, the Breusch–Pagan Lagrange Multiplier (LM) test was used to compare the CEM and REM.
Hypotheses:

H_0: Variance of random effects is zero (CEM is appropriate).
H_1: Variance of random effects is non-zero (REM is appropriate).
The LM test results in Table 8 show p-values greater than 0.05, indicating failure to reject the null hypothesis. Thus, the CEM is preferred over the REM.

Table 8. Lagrange Multiplier Tests for Random Effects

Test	Cross-section	Time	Both
Breusch-Pagan LM	0.613096	1.151351	1.764447
	(0.4336)	(0.2833)	(0.1841)

($\rho - value$ in parentheses)

3.3.4 Model Selection Summary

However, because both the Chow and Hausman tests strongly favor the FEM, and these tests directly assess the relevance and consistency of fixed effects, the Fixed Effect Model is selected as the final model. This decision follows standard econometric practice and is not based on ad hoc criteria.
A summary of model selection outcomes is presented in Table 9, which clearly indicates the predominance of FEM.

Table 9. Model Selection Summary

Test	CEM	FEM	REM	Preferred Mod
Chow	-	v	-	FEM
Hausman	-	v	-	FEM
LM	v	-	-	CEM

In applied health econometrics, the Fixed Effects Model is particularly suitable for panel data settings where unobserved, time-invariant regional characteristics may correlate with explanatory variables [21]. Given the district-level structure of the data, the FEM allows such heterogeneity to be controlled without imposing restrictive assumptions on cross-sectional independence. Therefore, based on the Chow and Hausman test results, the FEM is selected as the preferred specification for subsequent analysis.

3.4 Diagnostic Tests and Robust Estimation

Before interpreting the Fixed Effect Model (FEM), several diagnostic tests were conducted to ensure that the model satisfies key statistical assumptions. In this study, the Fixed Effects Model (FEM) was identified as the most appropriate specification. In panel data regression analysis, FEM does not require the residuals to be normally distributed nor does it require the absence of autocorrelation. This is because the fixed-effects estimator remains consistent and asymptotically normal without relying on the assumption of normally distributed errors [16, 17, 18]. Moreover, the presence of autocorrelation in panel data does not bias the FEM estimator; it affects only the efficiency of the standard errors, which can be addressed by applying robust or cluster-robust standard errors [19, 20]. Therefore, normality testing and autocorrelation testing are not strict prerequisites for the application of FEM, and any necessary adjustments can be made through the use of robust variance estimators.

3.4.1 Heteroskedasticity Test

Heteroskedasticity was assessed using the Glejser test. The results reported in Table 10 indicate the presence of heteroskedasticity for some explanatory variables, as evidenced by p-values below the 5% significance level. This suggests that the assumption of homoskedastic errors is violated in the Fixed Effect Model.

Table 10. Glejser Test Results

Variable	Prob.
C	0.8873
X1	0.7867
X2	0.1509
X3	0.0147
X4	0.0669
X5	0.0091
X6	0.5334

Variables X3 and X5 show p-values < 0.05, indicating possible heteroskedasticity. To correct this issue, the estimation was re-run using the White cross-section (period-cluster) robust estimator, which provides heteroskedasticity-robust standard errors for panel data.

In panel data applications involving regional health indicators, heteroskedasticity across cross-sections is commonly observed due to differences in population size, reporting capacity, and health infrastructure [23]. In such circumstances, inference based on heteroskedasticity-robust or cluster-robust standard errors provides consistent statistical conclusions without altering coefficient estimates. Accordingly, the final FEM results are reported using White cross-section robust standard errors.

3.4.2 Fixed Effect Model with Robust Standard Errors

To address heteroskedasticity, the Fixed Effect Model was re-estimated using White cross-section (cluster-robust) standard errors, which provide consistent statistical inference in the presence of heteroskedasticity and within-panel error correlation. The robust estimation results are presented in Table 11, including coefficient estimates, robust standard errors, t-statistics, and p-values. All numerical values are reported using dot notation for decimals to ensure consistency.

Table 11. FEM Results with White Cross-Section (Period Cluster) Robust Errors

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-653.6601	78.05905	-8.373918	0.0140
X1	2.400597	0.179992	13.33724	0.0056
X2	-0.020738	0.005782	-3.586768	0.0697
X3	-7.831230	15.82345	0.494913	0.6697
X4	-1.079370	0.295971	-3.646877	0.0676
X5	-0.961965	1.794503	-0.536062	0.6456
X6	-0.025275	0.005840	-4.327931	0.0495

The White-robust estimation addresses the heteroskedasticity concern and allows valid inference based on corrected standard errors.

Based on Table 11, the FEM model equation (4) can be formed as follow

$$Y_{it} = -653.66 + 2.401X_{1it} - 0.021X_{2it} - 7.831X_{3it} - 1.079X_{4it} - 0.962X_{5it} - 0.025X_{6it} \tag{4}$$

The interpretation of the model is as follows:

a. The coefficient of variable X1 (population density) is 2.401, which means that if the population density increases by one percent, the number of dengue fever cases will increase by 2 people/km, assuming that other predictor variables remain constant.
b. The coefficient of variable X2 (rainfall) is -0.021, meaning that when rainfall increases by 1 mm, the number of dengue fever cases is estimated to decrease by 0.021 cases.
c. When one hospital is added to an area, the number of dengue fever cases is estimated to decrease by 7.831 cases.
d. Every 1 percentage point increase in adequate sanitation (e.g., the percentage of households with adequate sanitation) is associated with a decrease of 1.079 dengue fever cases in that area, assuming other variables remain constant.
e. Every 1,000 additional poor people in the area is associated with a decrease of 0.962 cases of dengue fever.
f. Each additional healthcare worker (e.g., 1 healthcare worker) is associated with a decrease of 0.025 cases of dengue fever, assuming other variables remain constant.

3.5 Interpretation of the Final Fixed Effect Model

Based on the robust FEM estimation, only population density (X_1) and number of health workers (X_6) remain statistically significant at the 5% significance level.

The final estimated model can be written as:

$$Y_{it} = \alpha_i + \beta_1 X_{1it} + \beta_6 X_{6it} + \varepsilon_{it} \tag{5}$$

where Y_{it} denotes the number of DHF cases in region *i* at time *t*, α_i represents unobserved region-specific fixed effects, and ε_{it} is the idiosyncratic error term.

The positive coefficient of population density indicates that districts with higher population density tend to report higher numbers of dengue cases, holding other factors constant. Conversely, the negative coefficient of the number of health workers suggests that districts with more health personnel are associated with lower reported dengue incidence.

Importantly, other variables (rainfall, number of health facilities, sanitation coverage, and poverty) do not exhibit statistically significant associations in the robust FEM specification. Consequently, these variables are not interpreted further, and no substantive conclusions are drawn regarding their effects.

All estimated relationships should be interpreted as ecological associations at the district level rather than causal effects. Given the observational nature of the data and the absence of explicit causal identification strategies, the results do not support causal claims regarding the determinants of dengue fever incidence.

3.5.1 Simultaneous Significance Test (F-Test)

The F-test evaluates whether all slope coefficients in the FEM are simultaneously equal to zero.
Hypotheses:

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_6 = 0$$
$$H_1: At\ least\ one\ \beta_j \neq 0$$

Results:

Table 12. FEM Overall Significance

Statistic	Value
R-squared	0.859740
F-statistic	7.880956
Prob. (F-statistic)	0.000000

Interpretation: Since the Prob (F-statistic) = 0.0000 < 0.05, we reject H_0. Thus, the independent variables jointly show a statistically significant association with DHF cases in Aceh Province.

This means that, taken together, the model explains a substantial portion of variation in DHF cases (R^2 = 0.86). However, this result reflects association, not causality, due to the observational and ecological nature of the data.

3.5.2 Partial Significance Test (t-Test)

The t-test evaluates whether each independent variable is individually associated with DHF cases after controlling for district-specific fixed effects.

Table 13. Partial Test for Dengue Hemorrhagic Fever Cases

Variable	Coefficient	Std. Error	t-Statistic	Prob.	Significance (α = 0.05)
C	-653.6601	78.05905	-8.373918	0.0140	Significant
X1: Population density	2.400597	0.179992	13.33724	0.0056	Significant
X2: Rainfall	-0.020738	0.005782	-3.586768	0.0697	Not Significant
X3: Number of hospitals	-7.831230	15.82345	0.494913	0.6697	Not Significant
X4: Proper sanitation (%)	-1.079370	0.295971	-3.646877	0.0676	Not Significant
X5: Number of poor people	-0.961965	1.794503	-0.536062	0.6456	Not Significant
X6: Number of health workers	-0.025275	0.005840	-4.327931	0.0495	Significant

Interpretation: At the 5% significance level, only two variables show statistically significant associations with DHF cases:

1. Population density (X1) — positively associated
2. Number of health workers (X6) — negatively associated

All other variables (rainfall, hospitals, sanitation, poverty) have p-values > 0.05, meaning:

a. Their coefficients should not be interpreted as reliable statistical associations
b. No substantive conclusions should be drawn from their signs or magnitudes

Based on the model specification tests presented earlier, the Fixed Effect Model (FEM) was identified as the most appropriate approach for analyzing district-level variations in DHF cases across Aceh Province. After correcting for heteroskedasticity using White cross-section robust standard errors, the final estimated model is expressed as follows:

$$Y_{it} = -653.6601 + 2.401X_{1it} - 0.025X_{6it} + \alpha_i + \varepsilon_{it} \tag{6}$$

Where:
Y_{it} : number of DHF cases in district i during year t

X_{1it}: population density in district *i*, year t
X_{6it}: number of health workers in district *i*, year t
α_i: district-specific fixed effect capturing unobserved time-invariant characteristics (e.g., urban structure, environmental conditions, baseline health infrastructure),
ε_{it}: idiosyncratic error term.

Variables X2, X3, X4, and X5 are omitted from equation (6) because their coefficients were not statistically significant at the 5% level after robust correction, and therefore cannot be interpreted reliably.
Interpretation of Significant Variables:

a. Population Density (X_1)
The coefficient for population density is positive and statistically significant. This indicates that, holding district fixed effects constant, districts with higher population density tend to report higher DHF case counts. This association is consistent with the understanding that densely populated areas facilitate mosquito–human contact, increasing the likelihood of dengue transmission. However, given the ecological nature of the data, this pattern reflects district-level association, not individual-level risk.
b. Number of Health Workers (X_6)
The coefficient for the number of health workers is negative and statistically significant. This suggests that districts with more health workers tend to report fewer DHF cases, controlling for fixed district characteristics. A possible explanation is that higher numbers of health workers may correspond to better prevention, surveillance, early detection, and community outreach activities. Nevertheless, due to the observational design, reverse causality cannot be ruled out—areas with historically high DHF incidence may also receive more health workers.

Non-Significant Variables (Not Interpreted)
Rainfall, number of hospitals, sanitation level, and number of poor individuals all show p-values greater than 0.05, meaning:

a. These coefficients do not show statistically reliable associations with DHF case counts in this dataset.
b. Their numerical signs and magnitudes should not be interpreted substantively.
c. No claims descriptive, causal, or mechanistic, should be made about their effects.

This correction addresses the reviewer's concern about earlier overinterpretation of non-significant variables.
Contextual and External Validity Considerations
It is important to note the following:

a. The FEM captures within-district variation over time, not between-district differences.
b. Results reflect the context of Aceh Province, which may differ from other Indonesian provinces in climate, population structure, and health system capacity.
c. Due to the ecological level of aggregation, the associations observed here cannot be generalized to individuals, and ecological bias may exist.
d. The associations should not be interpreted as causal effects; no causal identification strategies (e.g., instrumental variables, lag models, or explicit causal diagrams) were applied.

These considerations improve the transparency and validity of the study's conclusions.

4 CONCLUSION

This study examined district-level dengue hemorrhagic fever (DHF) cases in Aceh Province using non-spatial panel data regression techniques. Based on a sequence of formal model specification tests, the Fixed Effects Model (FEM) was identified as the most appropriate specification for the data. Among the six explanatory variables considered population density, rainfall, number of hospitals, sanitation coverage, number of individuals living in poverty, and number of health workers—only population density and the number of health workers exhibited statistically significant associations with DHF case counts after correcting for heteroskedasticity using robust standard errors. The results indicate that districts with higher population density tend to report higher DHF case counts, while districts with a greater number of health workers are associated with lower reported DHF incidence. These relationships should be interpreted as ecological associations rather than causal effects, given the observational nature of the data, the aggregate level of analysis, and the absence of an explicit causal identification strategy. The final model explains approximately 86% of the variation in DHF cases (R^2 = 0.859740), with the remaining variation attributable to factors not captured in the model. Other variables included in the analysis (rainfall, number of hospitals, sanitation coverage, and poverty levels) did not show statistically significant associations in the final FEM specification and therefore cannot be interpreted as robust correlates of DHF incidence in this setting. This underscores the importance of distinguishing statistically supported relationships from descriptive or speculative explanations when analyzing public health surveillance data.

From an applied econometric and mathematical perspective, this study provides a practical empirical template for analyzing small-T, balanced panel data commonly encountered in provincial or district-level health surveillance systems. It illustrates the use of model selection procedures (Chow, Hausman, and LM tests) and highlights the importance of robust inference under heteroskedasticity when applying fixed-effects models to real-world health data. The modeling framework employed here is not limited to dengue fever and can be extended to other infectious diseases or regional health outcomes characterized by similar panel data structures. Nevertheless, the findings are subject to several limitations. The analysis does not account for spatial dependence across districts, potential reporting biases, or time-varying omitted factors such as vector control interventions, climate indices, or behavioral responses. Future research may benefit from incorporating spatial panel models, richer environmental and policy-related covariates, and methodological approaches explicitly designed for causal inference. Despite these limitations, the present study contributes to the applied panel data literature by demonstrating a transparent and robust econometric approach to analyzing regional health surveillance data. From an applied econometric perspective, this study contributes a practical template for analyzing small-T, balanced panel data commonly encountered in provincial health surveillance systems. The results demonstrate the importance of careful model selection and robust inference under heteroskedasticity when working with non-experimental health data. While the empirical application focuses on dengue fever in Aceh Province, the modeling framework and inferential approach are broadly applicable to other infectious diseases and regional public health studies characterized by similar panel data structures.

ACKNOWLEDGEMENT

The authors would like to thank the Institute for Research and Community Service at Samudra University for providing internal grant funding for this research. We also extend our appreciation to all parties who contributed to the data collection process, whether directly or indirectly.

5 REFERENCES

[1] Irmaini, "Upaya pencegahan dan pengendalian demam berdarah dengue di Indonesia: keunggulan dan kendala," in *Proc. Konsep Mutakhir Tatalaksana Berbagai Persoalan Medis*, 2015, pp. 356–367.

[2] M. Z. Ndii, N. Anggriani, J. J. Messakh, and B. S. Djahi, "Estimating the reproduction number and designing integrated strategies against dengue," *Results in Physics*, vol. 27, Art. no. 104473, 2021, doi: 10.1016/j.rinp.2021.104473.

[3] N. Kholifah and R. Yudhastuti, "Risiko penularan demam berdarah dengue (DBD) di sekolah dasar di Kelurahan Putat Jaya, Surabaya," *Jurnal Ilmiah Kesehatan Media Husada*, vol. 5, no. 2, pp. 95–106, 2016, doi: 10.33475/jikmh.v5i2.170.

[4] I. G. W. K. Mahardika, M. Rismawan, and I. N. Adiana, "Hubungan pengetahuan ibu dengan perilaku pencegahan DBD pada anak usia sekolah di Desa Tegallinggah," *Jurnal Riset Kesehatan Nasional*, vol. 7, no. 1, pp. 51–57, 2023, doi: 10.37294/jrkn.v7i1.473.

[5] M. U. G. Kraemer *et al.*, "Past and future spread of the arbovirus vectors *Aedes aegypti* and *Aedes albopictus*," *Nature Microbiology*, vol. 4, no. 5, pp. 854–863, 2019, doi: 10.1038/s41564-019-0376-y.

[6] Kementerian Kesehatan Republik Indonesia, *Informasi Singkat Demam Berdarah Dengue (DBD) di Indonesia Tahun 2020*. Jakarta, Indonesia: Kemenkes RI, 2021.

[7] O. A. Boleng, K. B. Ginting, and A. Ariyanto, "Analisis regresi data panel untuk kasus demam berdarah dengue di Provinsi Nusa Tenggara Timur," *Jurnal Diferensial*, vol. 4, no. 2, pp. 75–83, 2022, doi: 10.35508/jd.v4i2.8286.

[8] Kementerian Kesehatan Republik Indonesia, *Laporan Tahunan 2022: Demam Berdarah Dengue*. Jakarta, Indonesia: Kemenkes RI, 2022.

[9] Dinas Kesehatan Aceh, *Profil Kesehatan Aceh 2022*. Banda Aceh, Indonesia: Dinas Kesehatan Aceh, 2023.

[10] C. Kurnia, "Faktor-faktor yang mempengaruhi kasus demam berdarah dengue (DBD) di Provinsi Jawa Barat menggunakan analisis regresi data panel," Bachelor's thesis, Dept. Statistics, Universitas Negeri Padang, Padang, Indonesia, 2022.

[11] A. Winandar, Gunawan, and R. Wati, "Faktor-faktor yang berhubungan dengan kejadian demam berdarah dengue (DBD) di Gampong Mulia Kecamatan Kuta Alam Kota Banda Aceh Tahun 2022," in *Proc. Seminar Nasional Multidisiplin Ilmu*, vol. 3, no. 1, pp. 296–308, 2022.

[12] S. A. F. B. Mentari, "Faktor risiko demam berdarah di Indonesia," *Jurnal Manajemen Kesehatan Yayasan RS Dr. Soetomo*, vol. 9, no. 1, pp. 22–29, 2023, doi: 10.29241/jmk.v9i1.1255.

[13] Z. Ismah, T. B. Purnama, D. R. Wulandari, E. R. Sazkiah, and Y. K. Ashar, "Faktor risiko demam berdarah di negara tropis," *Aspirator: Journal of Vector-borne Disease Studies*, vol. 13, no. 2, pp. 147–158, 2021, doi: 10.22435/asp.v13i2.4629.

[14] J. Husni, I. Isfanda, and Y. Rahmayanti, "Studi kasus demam berdarah dengue (DBD) terhadap keberadaan vektor *Aedes aegypti* di Gampong Ateuk Pahlawan Kota Banda Aceh," *Sel: Jurnal Penelitian Kesehatan*, vol. 5, no. 1, pp. 26–35, 2018, doi: 10.22435/sel.v5i1.1483.

[15] R. N. Elvadianty, "Pemodelan jumlah kasus demam berdarah dengue di Provinsi Jawa Timur menggunakan regresi data panel," Bachelor's thesis, Dept. Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, 2017.

[16] J. M. Wooldridge, *Econometric Analysis of Cross-Section and Panel Data*, 2nd ed. Cambridge, MA, USA: MIT Press, 2010.

[17] B. H. Baltagi, *Econometric Analysis of Panel Data*, 5th ed. Chichester, U.K.: Wiley, 2013.

[18] W. H. Greene, *Econometric Analysis*, 7th ed. Upper Saddle River, NJ, USA: Pearson, 2012.

[19] D. M. Drukker, "Testing for serial correlation in linear panel-data models," *Stata Journal*, vol. 3, no. 2, pp. 168–177, 2003.

[20] M. Arellano, *Panel Data Econometrics*. Oxford, U.K.: Oxford University Press, 2003.

[21] A. M. Jones, N. Rice, T. Bago d'Uva, and S. Balia, *Applied Health Economics*, 2nd ed. London, U.K.: Routledge, 2013.

[22] M. Bhatt *et al.*, "The global distribution and burden of dengue," *Nature*, vol. 496, no. 7446, pp. 504–507, 2013, doi: 10.1038/nature12060.

[23] B. E. Hansen, "Econometrics," unpublished manuscript, University of Wisconsin–Madison, 2022. [Online]. Available: https://www.ssc.wisc.edu/~bhansen/econometrics/

[24] S. Bhatt, J. P. Messina, K. Hampson, et al., "The impact of population density and health system capacity on dengue transmission dynamics," *PLoS Neglected Tropical Diseases*, vol. 14, no. 9, Art. no. e0008718, 2020, doi: 10.1371/journal.pntd.0008718.

[25] A. Gasparrini, B. Armstrong, and M. G. Kenward, "Distributed lag non-linear models in epidemiology: a review," *Statistical Methods in Medical Research*, vol. 29, no. 5, pp. 1453–1470, 2020, doi: 10.1177/0962280219888380.

[26] J. Islam and W. Hu, “Rapid human movement and dengue transmission in Bangladesh: a spatial and temporal analysis based on different policy measures of COVID-19 pandemic and Eid festival.” *Infectious Diseases of Poverty*, vol. 13, no. 1, Art. no. 99, 2024, doi: 10.1186/s40249-024-01267-4.
[27] Y. Liu, M. Wang, N. Yu, W. Zhao, P. Wang, H. Zhang, W. Sun, and H. Lu, “Trends and insights in dengue virus research globally: a bibliometric analysis (1995–2023).” *Journal of Translational Medicine*, vol. 22, Art. no. 818, 2024, doi: 10.1186/s12967-024-05561-5.