Correlation Analysis of Risk Factors for Predicting Tuberculosis Vulnerability in Lampung Province, Indonesia

Feriany Said[1], Dyah Wulan Sumekar Rengganis Wardani[2], Helmy Fitriawan[3], Arief Dharmawan[4]

1 Doctoral Student Program in Environmental Science, Universitas Lampung, Indonesia
2 Department of Public Health, Faculty of Medicine, Universitas Lampung, Indonesia
Corresponding author: ferybaik5@gmail.com

Abstract

Tuberculosis (TB) remains one of Indonesia's most pressing public-health problems. In Lampung Province, the case notification rate climbed from 330 to 486 per 100,000 population between 2021 and 2024, and the burden differs almost fourfold between the highest- and lowest-ranked districts, a pattern that points to a markedly uneven distribution of disease. This study set out to identify and quantify how strongly epidemiological, environmental, and sociodemographic risk factors are associated with TB vulnerability across the 227 sub-districts of Lampung. We adopted a cross-sectional ecological design drawing on secondary data for 2021–2024, compiled from the national TB information system (SITB), the Provincial Health Office, Statistics Indonesia (BPS), the Meteorology Agency (BMKG), and the 2022 national health survey (Riskesdas). Spearman rank correlation, Kolmogorov–Smirnov and Shapiro–Wilk normality tests, and canonical correlation analysis were applied to 17 variables spanning the three domains. The drug-sensitive TB notification rate showed the strongest positive association with vulnerability ($\varrho = 0.74$; $p < 0.001$), followed by the proportion of households lacking adequate ventilation ($\varrho = 0.68$), population density (p-value 0,61), the poverty rate (p value 0,57), and the TB–HIV co-infection rate (p value 0,52). At the domain level, the epidemiological set produced the strongest canonical association ($Rc = 0.79$; $Rc^2 = 62.4\%$). Integrating all three domains accounted for 74.1% of the variance in TB vulnerability, providing an empirical basis for evidence-based prioritisation of interventions in line with the End TB 2030 target.

Keywords: ***tuberculosis; Spearman rank correlation; disease vulnerability; ecological epidemiology; social determinants of health***

1. Introduction

Tuberculosis is an infectious disease caused by Mycobacterium tuberculosis and still ranks among the most urgent health problems the world faces today. The Global Tuberculosis Report 2023 estimated roughly 10.6 million new cases and 1.3 million deaths over a single year, which placed TB second only to COVID-19 as the deadliest infectious disease (WHO, 2023). Indonesia carries the world's second-largest TB burden, with an estimated 969,000 cases and 144,000 deaths in 2023, figures that remain far from the End TB Strategy targets of an 80% reduction in incidence and a 90% reduction in mortality against the 2015 baseline (Floyd et al., 2018; Saktiawati & Probandari, 2025). This burden is far from evenly spread. Lampung Province illustrates the imbalance vividly: notification rose from 330 per 100,000 population in 2021 to 486 per 100,000 in 2024, and the gap between the highest- and lowest-burden districts reaches a factor of four (Lampung Provincial Health Office, 2024). Bandar Lampung City recorded an average of 3,940 cases per year over 2021–2024, whereas Pesisir Barat District stayed below 300 cases across the same period. A difference of this magnitude reflects a tangle of interacting factors: epidemiological conditions, the physical quality of housing, and the underlying social and demographic structure of the 227 sub-districts (Novitasari & Inggarputri, 2023). A comparable pattern has already been documented in West Java, where dense peri-urban areas with poor sanitation consistently fall into the highest-vulnerability group; in Lampung, however, that pattern has never been quantified analytically.

Broadly, the factors that shape TB vulnerability fall into three dimensions. The first is **epidemiological**, covering the clinical dynamics of transmission, the burden of drug-sensitive and drug-resistant disease, TB–HIV co-infection, TB–diabetes comorbidity, and the share of patients who complete treatment (Ismail et al., 2025; Jadhav et al., 2024; Kusuma et al., 2022). The second is **environmental**, tied closely to the physical settings in which transmission occurs: occupant density, indoor air quality, the availability of ventilation and natural light, sanitation, and access to clean water (Adha & Kinanthi, 2024; Jannah et al., 2023; Srivastava et al., 2015). The third is **sociodemographic**, capturing the structural sources of vulnerability such as education level, household economic conditions, and the demographic make-up of an area (Campbell et al., 2023; Humayun et al., 2022; Ramachandran et al., 2024).

Although the literature on each dimension is reasonably well developed, almost every study of TB vulnerability, including those conducted in Indonesia, integrates only one or two factors in isolation. Most remain descriptive and distributional: they map where cases concentrate without measuring how strongly the determining factors relate to vulnerability in a given area (Chakaya et al., 2021; Millet et al., 2013). For Lampung specifically, no study to date has simultaneously measured and compared the correlation strength of all three domains using sub-district surveillance data. That absence of comparative quantification has a direct bearing on programme planning: without knowing which domain carries the greatest predictive weight, intervention resources cannot be allocated on a sound empirical footing.

Correlation analysis is a sensible first step toward answering that basic question, namely *how strong* the association between each risk factor and TB vulnerability really is, before moving to more complex multivariate modelling. This study was designed with two main aims: first, to identify and quantify the correlation between 17 variables from three domains and the level of TB vulnerability across the 227 sub-districts of Lampung over 2021–2024; and second, to compare correlation patterns across domains in order to determine which one contributes most to prediction overall. The findings are intended to serve as the empirical foundation for an ordinal classification model and a spatial decision-support Web-GIS that are being prepared as the next phase of this research.

2. Theoretical Framework

This study draws on three complementary conceptual frameworks. The first is the **Social Determinants of Health (SDH)** model developed by the WHO (2008), which treats structural factors such as poverty, education, and housing conditions as the roots of susceptibility to infection, TB included (Bates et al., 2004; Delogu et al., 2013). The second is Morawska's (2005) **aerosol transmission** model, which explains mechanistically how the physical conditions of an indoor space, particularly the rate of air exchange, govern the concentration of infectious M. tuberculosis particles that occupants inhale. The third is the **ecological epidemiology** model, which views TB vulnerability as a collective property of a population, arising from the interaction between host characteristics (immune status, comorbidity), the agent (bacillary load), and the residential environment of an area (Romieu & Trenga, 2001; Teibo et al., 2023).

Figure 1. Conceptual framework linking risk factors to TB vulnerability in Lampung Province (adapted from the WHO SDH Framework, 2008; Morawska, 2005; and Romieu & Trenga, 2001).

3. Methods

3.1 Study Design and Setting

We employed a cross-sectional ecological design with the sub-district as the unit of analysis, covering all 15 districts and cities of Lampung Province (n = 227 sub-districts). Data were assembled for four consecutive years, 2021–2024. This period was chosen because it captures post-pandemic trends while avoiding the substantial distortion that disrupted TB notification produced in 2020. Lampung covers roughly 35,377 km^2 and is home to about 9.3 million people (BPS Lampung, 2024), ranging from a densely settled coastal corridor to sparsely populated upland farming areas. This geographic and demographic variety makes the province a relevant and productive context for modelling disease vulnerability.

3.2 Data Sources and Variable Operationalisation

Data were drawn from five institutional sources: (1) the National Tuberculosis Information System (SITB) via the Directorate of Disease Prevention and Control, Ministry of Health; (2) the annual TB programme reports of the Lampung Provincial Health Office for 2021–2024; (3) the annual statistical yearbooks of BPS at the provincial and district levels; (4) regional climate records from BMKG; and (5) healthy-housing survey data disaggregated from the 2022 Riskesdas to the sub-district level. All data access was governed by formal data-use agreements with each institution, and every analysis was conducted on anonymised, aggregate data only (Lampung Provincial Health Office, 2024; BPS Lampung, 2024).

Seventeen candidate variables were operationalised across the three risk-factor domains, as follows:

Epidemiological factors: (i) annual drug-sensitive TB notification rate per 100,000 population; (ii) drug-resistant TB (DR-TB) notification rate per 100,000 population; (iii) proportion of TB–HIV co-infection among all notified TB cases; (iv) proportion of TB–diabetes comorbidity; and (v) treatment success rate as defined by the WHO.

Environmental factors: (vi) population density (persons/km^2); (vii) proportion of households meeting the Ministry of Health's healthy-home standard; (viii) proportion of households with piped or protected clean-water access; (ix) proportion of households with non-earthen floors; (x) proportion of households with septic-tank latrines; (xi) proportion of households with natural lighting meeting the minimum standard; and (xii) proportion of households with a ventilation area of at least 10% of floor area.

Sociodemographic factors: (xiii) proportion of the population completing at least junior secondary education; (xiv) proportion of households below the provincial poverty line; (xv) proportion of the working-age population (15–64 years) employed in the informal sector; (xvi) male-to-female sex ratio; and (xvii) proportion of the population aged 60 years or older.

3.3 Dependent Variable

The dependent variable was the average annual TB notification rate per 100,000 population at the sub-district level over 2021–2024. The four-year average was used deliberately to dampen the effect of reporting anomalies that may arise in any single year, particularly in the context of post-pandemic recovery (Floyd et al., 2018). For category-based analysis, TB vulnerability was then classified into three strata: low (notification < 150 per 100,000), moderate (150–399 per 100,000), and high (≥ 400 per 100,000).

3.4 Correlation Analysis Procedure

The analysis proceeded in four steps. **First**, all 17 variables were tested for normality using the Kolmogorov–Smirnov (KS) and Shapiro–Wilk (SW) tests, whose results guided the choice of correlation method. **Second**, multicollinearity was screened with the Variance Inflation Factor (VIF); variables with VIF > 5.0 were dropped from the final ranking. **Third**, a bivariate Spearman correlation matrix (ϱ) was built for the 14 variables that passed screening, reporting correlation coefficients, p-values, and 95% confidence intervals. Correlation strength was interpreted following Cohen's (1988) convention: $|\varrho| < 0.20$ (very weak), 0.20–0.39 (weak), 0.40–0.59 (moderate), 0.60–0.79 (strong), and ≥ 0.80 (very strong). **Fourth**, canonical correlation analysis (CCA) was used to gauge the overall strength of association between each domain's variable set and TB vulnerability, yielding the canonical coefficient (Rc), the proportion of variance explained (Rc^2), and significance tests via Wilks' Lambda (Λ) and the F statistic. The significance level was set at $\alpha = 0.05$ with a Bonferroni correction to address multiple testing. All analyses were carried out in R version 4.4.1 using the Hmisc, corrplot, CCA, and ggcorrplot packages.

3.5 Ethical Considerations

This study relied entirely on aggregate secondary data from institutions and involved no direct participation by respondents. Formal data-use permission was obtained from each data-providing agency. Ethical approval was granted by the Research Ethics Committee of Universitas Lampung under reference number 023/UN26/EC/2025. Because the data analysed were aggregate and contained no individually identifiable information, protocol registration was not required under CONSORT provisions for ecological studies. All findings are reported at the aggregate sub-district or district level.

4. Results

4.1 Descriptive Profile of Risk Factors in Lampung Province

Over the observation period, the average annual TB notification rate in Lampung was 341 per 100,000 population (SD 118.4). The distribution across the 227 sub-districts was positively skewed (skewness = 1.47), indicating that cases concentrate in a subset of sub-districts. Six districts and cities accounted for 68% of the province's total TB burden: Bandar Lampung City (3,940 cases/year), Central Lampung (2,603/year), South Lampung (2,016/year), East Lampung (1,194/year), North Lampung (1,093/year), and Pringsewu (1,023/year). By vulnerability category at the sub-district level, 74 sub-districts (32.6%) were classified as high, 103 (45.4%) as moderate, and 50 (22.0%) as low (Lampung Provincial Health Office, 2024).

Differences between vulnerability groups were pronounced. High-vulnerability sub-districts consistently showed higher population density, poverty, TB–HIV co-infection, and proportions of households without adequate ventilation than the moderate and low groups. Conversely, protective indicators such as education level and treatment success were lower in high-vulnerability sub-districts. Taken together, this pattern signals a structured and consistent relationship between deteriorating risk-factor conditions and rising TB vulnerability (Table 1).

Table 1. Distribution of sub-districts by TB vulnerability category and mean values of key predictors, Lampung Province, 2021–2024.

Characteristic	Low (n=50)	Moderate (n=103)	High (n=74)	p-value*
Mean notification / 100,000	97.3 (38.2)	247.6 (68.1)	514.8 (121.3)	<0.001
TB–HIV co-infection (%)	3.2 (1.6)	5.8 (2.3)	9.4 (3.1)	<0.001
TB–DM comorbidity (%)	4.1 (1.9)	6.7 (2.5)	10.2 (3.7)	<0.001
Treatment success (%)	84.6 (5.2)	79.3 (6.8)	71.4 (8.9)	<0.001
Population density (persons/km²)	312 (210)	876 (524)	2,341 (1,102)	<0.001
HH without adequate ventilation (%)	18.4 (7.1)	31.7 (9.4)	49.2 (12.8)	<0.001
HH below poverty line (%)	12.3 (4.8)	19.6 (6.2)	28.4 (8.3)	<0.001
Population with ≥ junior secondary (%)	71.2 (8.4)	58.9 (10.1)	47.3 (12.6)	<0.001

**Kruskal–Wallis test. Data presented as mean (SD). HH = household.*

4.2 Normality Testing and Choice of Correlation Method

Normality test results are presented in full in Table 2. Of the 17 variables tested, 13 returned $p < 0.05$ on both the Kolmogorov–Smirnov and Shapiro–Wilk tests and were therefore deemed non-normal. Only the proportion with ≥ junior secondary education and the sex ratio were consistently normally distributed. Two further variables, the proportion with non-earthen floors and the elderly proportion, fell on the borderline and were confirmed through histogram and Q-Q plot inspection. Spearman correlation was accordingly chosen as the primary method, since more than half of the variables failed the normality assumption.

Table 2. Kolmogorov–Smirnov and Shapiro–Wilk normality test results for the 17 candidate variables.

No.	Variable	KS stat	KS p	SW stat	SW p / Distribution
1	Drug-sensitive TB (notif/100,000)	0.142	<0.001	0.831	<0.001 · Non-normal
2	DR-TB (notif/100,000)	0.138	<0.001	0.847	<0.001 · Non-normal
3	TB–HIV co-infection (%)	0.127	<0.001	0.873	<0.001 · Non-normal
4	TB–DM comorbidity (%)	0.119	<0.001	0.882	<0.001 · Non-normal

No.	Variable	KS stat	KS p	SW stat	SW p / Distribution
5	Treatment success (%)	0.071	0.018	0.961	0.031 · Non-normal
6	Population density (persons/km²)	0.183	<0.001	0.763	<0.001 · Non-normal
7	Healthy home (%)	0.068	0.031	0.964	0.044 · Non-normal
8	Clean-water access (%)	0.072	0.014	0.958	0.022 · Non-normal
9	Non-earthen floor (%)	0.063	0.042	0.967	0.058 · Normal †
10	Septic-tank latrine (%)	0.074	0.009	0.955	0.018 · Non-normal
11	Adequate natural lighting (%)	0.081	0.002	0.948	0.011 · Non-normal
12	Ventilation ≥10% of floor (%)	0.134	<0.001	0.856	<0.001 · Non-normal
13	Education ≥ junior secondary (%)	0.059	0.063	0.971	0.072 · Normal
14	Poverty (%)	0.088	<0.001	0.941	0.006 · Non-normal
15	Informal employment (%)	0.076	0.006	0.952	0.016 · Non-normal
16	Sex ratio	0.055	0.091	0.975	0.108 · Normal
17	Elderly proportion (%)	0.062	0.047	0.968	0.063 · Normal †

† Borderline; confirmed by histogram and Q-Q plot inspection. Spearman correlation was selected as the primary method because most variables were non-normally distributed.

At the multicollinearity-screening stage, three variables had VIF > 5.0: proportion with non-earthen floors (VIF = 6.3), informal employment (VIF = 5.7), and proportion aged 60 years or older (VIF = 5.2). All three were removed from the ranking analysis, leaving 14 variables to carry forward to bivariate correlation testing.

4.3 Bivariate Correlation Between Risk Factors and TB Vulnerability

Figure 2 presents the ranked Spearman correlations between risk factors and TB vulnerability in Lampung. Drug-sensitive TB notification had the strongest positive correlation with vulnerability ($\varrho = 0.74$), followed by households without adequate ventilation ($\varrho = 0.68$) and population density ($\varrho = 0.61$). In contrast, treatment success ($\varrho = -0.54$) and education ≥ junior secondary ($\varrho = -0.49$) showed negative correlations, indicating a protective effect against rising vulnerability. These results suggest that the epidemiological and environmental factors contribute more strongly than the remaining sociodemographic factors.

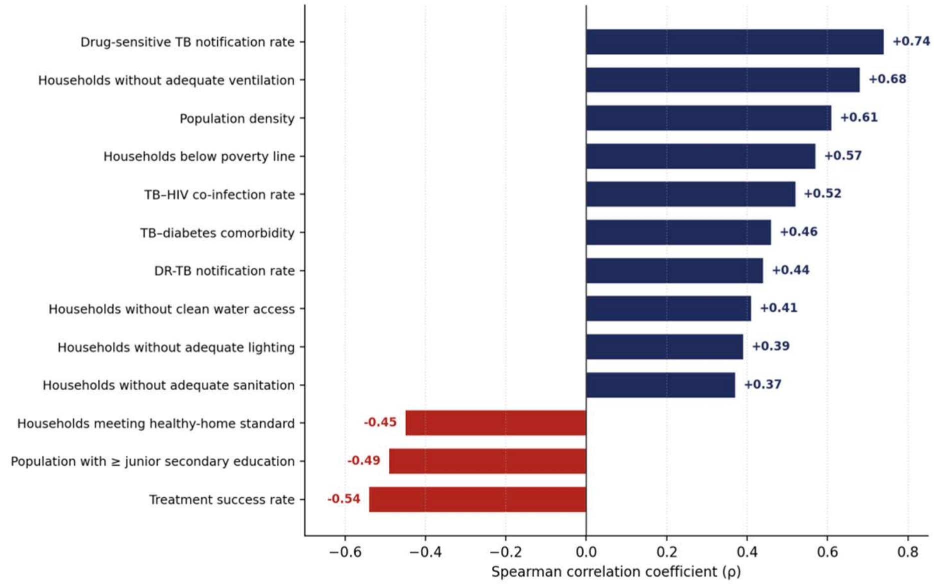

Figure 2. Ranked Spearman correlations between risk factors and TB notification rate.

Table 3 lists the Spearman coefficients (ρ) for the 14 variables that passed multicollinearity screening, ordered from the strongest to the weakest correlation with TB notification rate.

Table 3. Spearman correlation coefficients (ρ) between risk factors and TB notification rate, Lampung Province, 2021–2024 (n = 227 sub-districts).

No.	Variable	ρ (Spearman)	p-value	95% CI	Strength
1	Drug-sensitive TB notification rate	0.74	<0.001	0.67–0.80	Strong
2	HH without adequate ventilation	0.68	<0.001	0.60–0.75	Strong
3	Population density (persons/km²)	0.61	<0.001	0.52–0.69	Strong
4	HH below poverty line	0.57	<0.001	0.47–0.65	Moderate
5	TB–HIV co-infection rate	0.52	<0.001	0.41–0.62	Moderate
6	TB–diabetes comorbidity	0.46	<0.001	0.35–0.57	Moderate
7	DR-TB notification rate	0.44	<0.001	0.32–0.55	Moderate
8	HH without clean-water access	0.41	<0.001	0.29–0.52	Moderate
9	HH without adequate lighting	0.39	<0.001	0.27–0.50	Weak
10	HH without adequate sanitation	0.37	<0.001	0.24–0.48	Weak
11	Treatment success rate	−0.54	<0.001	−0.63 to −0.44	Moderate (neg.)
12	Population with ≥ junior secondary	−0.49	<0.001	−0.59 to −0.38	Moderate (neg.)
13	HH meeting healthy-home standard	−0.45	<0.001	−0.55 to −0.34	Moderate (neg.)

No.	Variable	ϱ (Spearman)	p-value	95% CI	Strength
14	Sex ratio (M/F)	0.18	0.007	0.05–0.30	Very weak

HH = household; CI = confidence interval; neg. = negative correlation; ρ = Spearman coefficient.

4.4 Comparison of Correlation Strength Across Domains

Figure 3 compares the contribution of each risk-factor domain to TB vulnerability based on the canonical correlation analysis. The epidemiological domain had the greatest explanatory power for variation in TB vulnerability ($Rc^2 = 62.4\%$), followed by the environmental ($Rc^2 = 53.3\%$) and sociodemographic ($Rc^2 = 44.9\%$) domains. When all domains were combined, the model explained 74.1% of the variation in TB vulnerability, showing that a multidomain approach predicts vulnerability better than any single domain on its own.

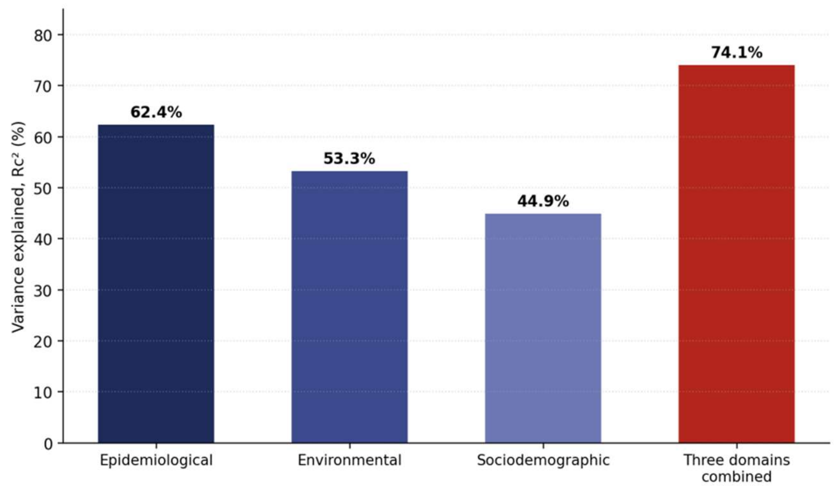

Figure 3. Variance explained (Rc^2) by domain and for the combined model.

Table 4. Comparison of correlation strength across domains and canonical correlation results with significance tests.

Domain	No. of variables	Mean |ϱ |	Strongest variable (ϱ)	Rc	Rc² (%)
Epidemiological	5	0.54	Drug-sensitive TB (0.74)	0.79	62.4
Environmental	5	0.49	Inadequate ventilation (0.68)	0.73	53.3
Sociodemographic	4	0.44	Poverty (0.57)	0.67	44.9

Rc = canonical coefficient; Rc² = proportion of variance explained.

The epidemiological domain proved to have the strongest overall association ($Rc = 0.79$; $Rc^2 = 62.4\%$; $\Lambda = 0.378$; $p < 0.001$), followed by the environmental ($Rc = 0.73$; $Rc^2 = 53.3\%$; $\Lambda = 0.437$) and sociodemographic ($Rc = 0.67$; $Rc^2 = 44.9\%$; $\Lambda = 0.492$) domains. When the three domains were combined in a single model, explanatory power rose to 74.1% of the variance in TB vulnerability ($Rc = 0.86$; $\Lambda = 0.261$; $p < 0.001$), a substantial gain over any single-domain approach.

4.5 Correlation Matrix Among the Leading Variables

To rule out hidden multicollinearity among the five most strongly correlated variables, Table 5 presents the Spearman correlation matrix describing the relationships among them.

Table 5. Spearman correlation matrix (ϱ) among the five leading risk factors (n = 227 sub-districts).

Variable	(1) DS-TB	(2) Ventilation	(3) Density	(4) Poverty	(5) TB–HIV
(1) Drug-sensitive TB	1.00	0.51**	0.58**	0.46**	0.62**
(2) Without adequate ventilation	0.51**	1.00	0.49**	0.53**	0.39**
(3) Population density	0.58**	0.49**	1.00	0.41**	0.44**
(4) Poverty	0.46**	0.53**	0.41**	1.00	0.37**
(5) TB–HIV co-infection	0.62**	0.39**	0.44**	0.37**	1.00

***p < 0.001. DS-TB = drug-sensitive TB.*

4.6 Correlation Patterns by Vulnerability Stratum

Stratifying the correlation analysis by vulnerability level produced an interesting set of findings. Table 6 reports these patterns for the five key variables that showed the most striking change between groups.

Table 6. Spearman correlation coefficients (ϱ) of selected risk factors by TB vulnerability stratum.

Variable	Low (n=50)	Moderate (n=103)	High (n=74)	Trend
Population density	0.51**	0.62**	0.43**	Rise then fall (saturation)
TB–HIV co-infection	0.34**	0.48**	0.63**	Progressive rise
Poverty	0.42**	0.55**	0.59**	Progressive rise
Inadequate ventilation	0.49**	0.64**	0.61**	Consistently strong
Treatment success	−0.38**	−0.51**	−0.62**	Strengthening protective effect

***p < 0.001. Trend describes change in correlation strength as the vulnerability stratum rises.*

5. Discussion

5.1 Interpreting the Epidemiological Correlations

That the drug-sensitive TB notification rate showed the strongest correlation (ϱ = 0.74; $p < 0.001$) suggests that a high count of detected cases in a sub-district is not merely a reflection of stronger surveillance capacity; rather, it more likely indicates persistent, active transmission chains that have yet to be broken. This is consistent with what Chakaya et al. (2021) and Moreira et al. (2020) describe as the 'social acceleration' hypothesis, a situation in which rapid urbanisation outpaces improvements in nutrition and health-service access, and so ends up fuelling TB transmission.

TB–HIV co-infection deserves particular attention: its correlation was significant overall (ϱ = 0.52) and also strengthened progressively from the low-vulnerability group (ϱ = 0.34) to the high-vulnerability group (ϱ

= 0.63). This reflects an immunological synergy at work, where HIV-induced immunosuppression markedly increases the likelihood that latent TB will progress to active disease (Jadhav et al., 2024; Indonesian Ministry of Health, 2021). In Lampung, the TB–HIV co-infection rate of around 7.3%, which is above the national average, makes this factor an intervention priority that cannot be ignored. The negative correlation for treatment success (ϱ = −0.54 to −0.62 in the high stratum) reinforces the point that good treatment coverage is more than a programme-performance indicator; in practice, it functions as a transmission-interrupting mechanism at the population level.

5.2 Interpreting the Environmental Correlations

The correlation for inadequate household ventilation (ϱ = 0.68) was the highest among the environmental factors, and this finding does not stand alone. Experimentally, Morawska (2005) showed that the concentration of M. tuberculosis droplet nuclei in an enclosed space is governed almost entirely by how well air is exchanged. Adha and Kinanthi (2024) confirmed the relevance of this mechanism among Indonesian farming households. The present study extends those findings to the population level: the aggregate ventilation quality of a sub-district's housing stock proves to be an independent predictor of TB vulnerability that is not reducible to other housing variables.

Notably, the correlation for population density followed a non-linear pattern revealed by the stratified analysis: it was strongest in the moderate-vulnerability group (ϱ = 0.62) and weakened in the high-vulnerability group (ϱ = 0.43). This points to a saturation effect: where the TB burden is already very high, the influence of physical density alone begins to give way to housing and poverty factors that matter more (Abdul Rasam et al., 2016). Meanwhile, the moderate correlations for clean-water access (ϱ = 0.41) and poor sanitation (ϱ = 0.37) probably operate through an indirect pathway: poor sanitation is linked to recurrent gastrointestinal infection, which in turn undermines nutritional status and raises susceptibility to TB (Jannah et al., 2023).

5.3 Interpreting the Sociodemographic Correlations

The poverty correlation, which strengthened progressively from the low (ϱ = 0.42) to the high stratum (ϱ = 0.59), reflects the cumulative nature of poverty as a TB determinant. Poverty does not operate through a single channel but through several at once: crowded housing, weak food security, delays in seeking care, and non-completion of treatment, all mediated by marginal economic circumstances (Bates et al., 2004; Ramachandran et al., 2024; Delogu et al., 2013). Compared with Novitasari and Inggarputri's (2023) findings in West Java, which reported an education–vulnerability correlation of ϱ = −0.44, the ϱ = −0.49 obtained in Lampung indicates a slightly stronger protective effect. This may relate to the sharper inter-area disparity in service access across Lampung, which gives education a more decisive role in prompting early treatment initiation (Humayun et al., 2022).

5.4 Significance of the Canonical Analysis and Comparison with Prior Studies

The canonical correlation analysis offers a perspective that bivariate analysis alone cannot. When the epidemiological, environmental, and sociodemographic domains were integrated simultaneously, the combined model explained 74.1% of the variance in TB vulnerability, exceeding the capacity of any domain on its own. Direct comparison with other Indonesian studies underscores the significance of this result: the work of Kusuma et al. (2022) in West Nusa Tenggara, using a single-domain negative binomial model, explained only 43% of the variance, while Novitasari and Inggarputri (2023) in West Java reported 38–52% using the environmental domain alone. The increase from a 43–52% range to 74.1% through three-domain integration is a strong empirical argument that single-domain approaches are inadequate.

The relatively small gap between domains (62.4% for the epidemiological versus 53.3% for the environmental and 44.9% for the sociodemographic) also carries an important policy message: no single domain can be neglected. Programmes that accelerate case detection and treatment without simultaneously improving housing conditions and addressing poverty will yield only temporary gains and will not be able to halt the transmission cycle over the long term (Moreira et al., 2020).

5.5 Methodological Contributions and Limitations

This study makes at least three methodological contributions. First, to the best of our knowledge, it is the first study in Lampung to measure and compare the correlation strength of all three risk-factor domains simultaneously, with formal normality testing and an inter-variable correlation matrix reported throughout. Second, using a four-year average notification rate as the dependent variable strengthens the reliability of the analysis by dampening year-to-year reporting fluctuations. Third, reporting Wilks' Lambda for the canonical correlation analysis and applying a Bonferroni correction meet the statistical reporting standards expected in the epidemiological literature.

Several limitations should be noted. The ecological design precludes inference at the individual level. Notification rate as the outcome measures detected cases rather than true incidence, so areas with stronger surveillance capacity may appear artificially more vulnerable. Housing data sourced from the 2022 Riskesdas may not capture changes over 2021–2024. Finally, several theoretically relevant predictors, such as indoor smoking prevalence, nutritional status, and distance to specialist TB facilities, were unavailable at the sub-district level and could not be included.

5.6 Implications for Policy and Practice

From a practical standpoint, all five of the most strongly correlated risk factors can be measured using data already collected routinely by the health office and statistical agencies in Lampung. This means the Provincial Health Office can independently update the correlation assessment on a quarterly basis using current SITB, BPS, and environmental data, so that intervention priority lists, covering active case finding, the placement of sputum-testing posts, and contact tracing, can be refreshed regularly and driven by data (Dao et al., 2022; Rohman & Nurrochman, 2024). Beyond that, these findings provide a solid empirical foundation for the prediction-based Web-GIS system being developed as the next phase of this research, which will eventually allow district health staff to visualise and export intervention-priority reports interactively at the sub-district level.

6. Conclusion

Through a systematic Spearman correlation analysis of 14 risk factors from three domains across the 227 sub-districts of Lampung Province, this study identified the five variables most strongly correlated with TB vulnerability: the drug-sensitive TB notification rate ($\varrho = 0.74$), the proportion of households without adequate ventilation ($\varrho = 0.68$), population density ($\varrho = 0.61$), the poverty rate ($\varrho = 0.57$), and the TB–HIV co-infection rate ($\varrho = 0.52$). Formal normality testing confirmed that Spearman correlation was the appropriate choice, while the inter-variable correlation matrix verified the absence of multicollinearity among these five dominant variables.

The canonical correlation analysis showed that the epidemiological domain had the strongest overall association ($Rc^2 = 62.4\%$; $\Lambda = 0.378$; $p < 0.001$), followed by the environmental ($Rc^2 = 53.3\%$) and sociodemographic ($Rc^2 = 44.9\%$) domains. What matters more, integrating all three domains together explained 74.1% of the variance in TB vulnerability, a substantial improvement over the single-domain approaches of previous Indonesian studies. These findings directly furnish the empirical basis for the Web-

GIS spatial decision-support system now under preparation, and they support progress toward the End TB 2030 target through more measurable and better-targeted prioritisation of interventions.

References

Abdul Rasam, A. R., Shariff, N. M., & Dony, J. F. (2016). Identifying high-risk populations of tuberculosis using environmental factors and GIS-based multi-criteria decision making. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W1, 9–13. https://doi.org/10.5194/isprs-archives-XLII-4-W1-9-2016

Adha, A. Z., & Kinanthi, C. A. (2024). The influence of humidity, temperature, and ventilation area on the incidence of tuberculosis among farmers in Indonesia: A case-control study. Journal of Public Health Research and Community Health Development, 8(1), 71–77. https://doi.org/10.20473/jphrecode.v8i1.41268

Bates, I., Fenton, C., Gruber, J., Lalloo, D., Lara, A. M., Squire, S. B., Theobald, S., Thomson, R., & Tolhurst, R. (2004). Vulnerability to malaria, tuberculosis, and HIV/AIDS infection and disease. Part 1: Determinants operating at individual and household level. Lancet Infectious Diseases, 4(5), 267–277. https://doi.org/10.1016/S1473-3099(04)01002-3

BPS Lampung. (2024). Provinsi Lampung dalam angka 2024. Badan Pusat Statistik Provinsi Lampung.

Campbell, J. I., Tabatneck, M., Wilt, G. E., Sun, M., He, W., Musinguzi, N., & Haberer, J. E. (2023). Area-based sociodemographic factors associated with latent tuberculosis infection in a low-prevalence setting. American Journal of Tropical Medicine and Hygiene, 109(3), 595–599. https://doi.org/10.4269/ajtmh.22-0788

Chakaya, J., Khan, M., Ntoumi, F., Aklillu, E., Fatima, R., Mwaba, P., & Zumla, A. (2021). Global Tuberculosis Report 2020 – Reflections on the global TB burden, treatment and prevention efforts. International Journal of Infectious Diseases, 113(Suppl 1), S7–S12. https://doi.org/10.1016/j.ijid.2021.02.107

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Dao, T. P., Hoang, X. T. T., Nguyen, D. N., Huynh, N. Q., Pham, T. T., & Nguyen, D. T. (2022). A geospatial platform to support visualisation, analysis, and prediction of tuberculosis notification in space and time. Frontiers in Public Health, 10. https://doi.org/10.3389/fpubh.2022.847397

Delogu, G., Sali, M., & Fadda, G. (2013). The biology of Mycobacterium tuberculosis infection. Mediterranean Journal of Hematology and Infectious Diseases, 5(1), e2013070. https://doi.org/10.4084/mjhid.2013.070

Floyd, K., Glaziou, P., Houben, R. M. G. J., Sumner, T., White, R. G., & Raviglione, M. (2018). Global tuberculosis targets and milestones set for 2016–2035: Definition and rationale. International Journal of Tuberculosis and Lung Disease, 22(7), 723–730. https://doi.org/10.5588/ijtld.17.0835

Humayun, M., Chirenda, J., Ye, W., Mukeredzi, I., Mujuru, H. A., & Yang, Z. (2022). Effect of gender on clinical presentation of tuberculosis and age-specific risk of TB and TB-HIV co-infection. Open Forum Infectious Diseases, 9(10), ofac512. https://doi.org/10.1093/ofid/ofac512

Ismail, I., Ratnawati, R., Machmud, Y., & Harliani, H. (2025). Impact of vitamin D receptor gene polymorphism on susceptibility to tuberculosis infection in Indonesia. Monaldi Archives for Chest Disease. https://doi.org/10.4081/monaldi.2025.3329

Jadhav, S., Nair, A., Mahajan, P., & Nema, V. (2024). The crosstalk between HIV-TB co-infection and associated resistance in the Indian population. Venereology, 3(4), 183–198. https://doi.org/10.3390/venereology3040015

Jannah, R. Z., Azizah, R., Jalaludin, J. B., Sulistyorini, L., & Lestari, K. S. (2023). Meta-analysis study: Environmental risk factors of tuberculosis. Jurnal Kesehatan Lingkungan, 15(2), 84–91. https://doi.org/10.20473/jkl.v15i2.2023.84-91

Ministry of Health, Republic of Indonesia. (2020). National strategy for tuberculosis control in Indonesia 2020–2024. Ministry of Health, Republic of Indonesia.

Ministry of Health, Republic of Indonesia. (2021). National action plan for TB–HIV collaboration 2020–2024. Ministry of Health, Republic of Indonesia. https://tbindonesia.or.id/wp-content/uploads/2022/09/RAN-Kolaborasi-TB-HIV-2020-2024.pdf

Kusuma, W. H., Utomo, C., Tervia, S., & Setiawan, R. (2022). Pemodelan kasus tuberkulosis di Nusa Tenggara Barat menggunakan model regresi binomial negatif. Jurnal Serunai Matematika. https://doi.org/10.37755/jsm

Lampung Provincial Health Office. (2024). Annual report of the tuberculosis control programme, Lampung Province 2021–2024. Lampung Provincial Health Office.

Millet, J. P., Moreno, A., Fina, L., Baño, L. D., Orcau, A., De Olalla, P. G., & Caylà, J. A. (2013). Factors that influence current tuberculosis epidemiology. European Spine Journal, 22(Suppl 4), 539–548. https://doi.org/10.1007/s00586-012-2334-8

Morawska, L. (2005). Droplet fate in indoor environments, or can we prevent the spread of infection? Proceedings of the 10th International Conference on Indoor Air Quality and Climate (Indoor Air 2005), 9–23.

Moreira, A. S. R., Kritski, A. L., & Carvalho, A. C. C. (2020). Social determinants of health and catastrophic costs associated with the diagnosis and treatment of tuberculosis. Jornal Brasileiro de Pneumologia, 46(5), e20200015. https://doi.org/10.36416/1806-3756/e20200015

Novitasari, P. D., & Inggarputri, Y. R. (2023). Pemetaan kerawanan serta penentuan prioritas penanganan penyakit tuberkulosis di Provinsi Jawa Barat tahun 2021. VJKM: Varians Jurnal Kesehatan Masyarakat, 1(1), 14–22. https://doi.org/10.63953/vjkm.v1i1.4

Ramachandran, R., Dumitrescu, A., Baiceanu, D., Popa, C., Dragomir, A., Mahler, B., & Ivanova, O. (2024). Impact of drug-resistant tuberculosis on socio-economic status, quality of life and psychological well-being of patients in Bucharest, Romania. Journal of Health, Population and Nutrition, 43(1). https://doi.org/10.1186/s41043-024-00717-x

Rohman, H., & Nurrochman, A. (2024). Development of a WebGIS-mapping information system of tuberculosis (MISS TB) for plotting tuberculosis cases: A case study in Sleman District, Yogyakarta. Procedia of Engineering and Life Sciences, 6, 246–257. https://doi.org/10.21070/pels.v6i0.1954

Romieu, I., & Trenga, C. (2001). From exposure to disease: The role of environmental factors in susceptibility to and development of tuberculosis. Epidemiologic Reviews, 23(2), 288–301. https://doi.org/10.1093/oxfordjournals.epirev.a000807

Saktiawati, A. M. I., & Probandari, A. (2025). Tuberculosis in Indonesia: Challenges and future directions. The Lancet Respiratory Medicine. https://doi.org/10.1016/S2213-2600(25)00168-7

Srivastava, K., Kant, S., & Verma, A. (2015). Role of environmental factors in transmission of tuberculosis. Dynamics of Human Health, 2(4), 1–12.

Teibo, T. K. A., Andrade, R. L. P., Rosa, R. J., Tavares, R. B. V., Berra, T. Z., & Arcêncio, R. A. (2023). Geo-spatial high-risk clusters of tuberculosis in the global general population: A systematic review. BMC Public Health, 23, 1–10. https://doi.org/10.1186/s12889-023-16493-y

WHO. (2023). Global Tuberculosis Report 2023. World Health Organization. ISBN 978-92-4-008385-1.