Journal of Public Health and Pharmacy
2775-4952 (Online - Elektronik)
Research Articles
Open Access

Unveiling Tuberculosis Dynamics in Indonesia for Effective Control and Prevention: A Panel Regression and Clustering Approach

Novi Reandy Sasmita[1*], Mutiara Syifa[2], Mhd Khairul[3], Latifa Rahayu[4], Zurnila Marli Kesuma[5], Selvi Mardalena[6], Rumaisa Kruba[7], Virasakdi Chongsuvivatwong[8], M. Ischaq Nabil Asshiddiqi[9]
[1234567]Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh, Indonesia
[8]Department of Epidemiology, Faculty of Medicine, Prince of Songkla University, Hat Yai, Thailand
[9]School of Nursing, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

***Corresponding Author:** E-mail: novireandys@usk.ac.id

ARTICLE INFO

Manuscript Received: 08 Apr, 2025
Revised: 03 Dec, 2025
Accepted: 06 Jan, 2026
Date of publication: 01 Jul, 2026
Volume: 6
Issue: 2
DOI: 10.56338/jphp.v6i2.7265

KEYWORDS

Tuberculosis;
Panel Data Regression;
Fuzzy Possibilistic C-Means (FPCM);
Health and Environmental Factors;
Indonesia

ABSTRACT

Introduction: Tuberculosis (TB) remains a critical global health challenge, with Indonesia ranking second in global TB burden. This study examines factors influencing TB incidence across Indonesian provinces and applies clustering to guide targeted interventions aligned with TB eradication goals by 2030. Specifically, these findings inform Indonesia's End TB 2030 roadmap by identifying provincial heterogeneity that necessitates differentiated resource allocation and strengthened health governance frameworks.
Methods: This ecological time-series study design analyzed data from 34 Indonesian provinces (2020–2022), including TB cases, healthcare services, HIV cases, smoking prevalence, food management places, and public facilities. Descriptive statistics summarized variable distribution, while panel data regression identified key factors using multicollinearity checks, model selection, and assumption testing. Fuzzy Possibilistic C-Means (FPCM) clustering grouped provinces based on similarity characteristics.
Results: TB cases rose from 10,351 in 2020 to 21,303 in 2022. This study underscores the multifaceted factors influencing TB incidence in Indonesia. Significant factors included healthcare services (β_1 = -8.37), HIV cases (β_2 = 13.76), smoking prevalence (β_3 = 905.32), food management places (β_4 = 1.62), and public facilities (β_5 = 1.11). This study proves that TB is not only influenced by health factors but also by non-health factors. Fuzzy clustering using the FPCM identified three clusters based on their possibilistic membership degrees: Cluster 3, with high HIV prevalence and public facilities, requiring urgent action; Cluster 2, needing improved healthcare and smoking reduction; and Cluster 1, with moderate challenges.
Conclusions: Health and environmental factors significantly influence TB incidence. Addressing cluster-specific needs, such as enhancing healthcare, reducing HIV and smoking prevalence, and improving public health standards, is essential for TB control. Future studies should expand variables and periods to deepen insights.

Publisher: Pusat Pengembangan Teknologi Informasi dan Jurnal Universitas Muhammadiyah Palu

INTRODUCTION

Tuberculosis (TB), caused by Mycobacterium tuberculosis, continues to be a major global health challenge and remains a leading cause of mortality, particularly in developing countries (1,2). As of 2023, TB reemerged as the deadliest disease caused by a single infectious agent, accounting for an estimated 1.25 million deaths globally (95% UI: 1.13–1.37 million). Over 10 million people develop TB annually, with cases increasing since 2021, highlighting the urgent need for action to end the epidemic by 2030. In 2023, 55% of TB cases were men, 33% were women, and 12% were children or adolescents. Newly diagnosed cases rose to 8.2 million in 2023, up from 7.5 million in 2022, reflecting a backlog of undiagnosed TB during the pandemic years (3,4).

These trends highlight the need to rely on open national TB surveillance datasets such as the US National Tuberculosis Surveillance System (NTSS) run by the CDC, the Tuberculosis Information System (SITB) run by the Indonesia Government, and the UK Health Security Agency (UKHSA) TB system by the United Kingdom Goverment. These systems help improve the monitoring, reporting, and quality of TB services. sing these publicly accessible sources also enables ethically sound analyses by allowing independent verification of case reporting and disease trends.

Indonesia ranks second globally in TB cases, contributing 10% of the total. The top contributors include India (26%), Indonesia (10%), China (6.8%), the Philippines (6.8%), and Pakistan (6.3%). Between 2020 and 2023, Indonesia, the Philippines, and Myanmar were the main drivers of the global rise in TB cases. Indonesia recorded 1,090,000 TB cases, including 30,000 Multi-Drug-Resistant TB (MDR-TB), 575 Extensively Drug-Resistant TB (XDR-TB), and the remaining cases classified as Drug-Susceptible TB (DS-TB) (3,5). TB is the second leading cause of death in Indonesia after HIV/AIDS, with 724,309 cases reported in 2022, marking a significant annual increase (6,7).

Some studies highlighted that TB is affected not only by health factors but also by non-health factors such as individual behavior and environmental conditions (8,9). Previous studies found links between TB incidence and population density (10). Other studies showed associations between TB and HIV, along with access to sanitation (11,12). Ningrum et al. (2022) identified the influence of food management facilities and health centers on TB, while other findings emphasized the role of public spaces and smoking behavior (13).

Panel data regression is widely used to analyze factors influencing TB. This method examines data across time and entities, uncovering temporal dynamics and accounting for individual-specific characteristics that remain constant over time (14,15). Combining cross-sectional and time-series data provides more robust insights into the relationships between variables. It helps identify trends, correlations, and causal effects, making it a valuable tool for understanding the multifaceted nature of TB incidence and management.

Additionally, previous studies have used clustering in various contexts. There are two main types of clustering: hard clustering, which assigns each data point to a single cluster, and fuzzy clustering, which allows data points to belong to multiple clusters with varying degrees of membership (16,17). Fuzzy clustering is more powerful than hard clustering for handling complex and uncertain data, as it accommodates overlapping cluster boundaries and better represents real-world scenarios with inherent variability. It is especially suitable for applications involving noisy, imprecise, or incomplete datasets (18,19).

Fuzzy clustering analysis is essential for grouping Indonesian provinces based on TB-related factors. One of the well-known methods in fuzzy clustering is Fuzzy C-Means (FCM). While FCM is widely used, its sensitivity to initial conditions and susceptibility to local minima limit its effectiveness with noisy data or outliers. To address these limitations, Fuzzy Possibilistic C-Means (FPCM) was developed. FPCM introduces a possibilistic parameter that enables each data point to have independent membership degrees, enhancing its robustness to noise and outliers. This improvement makes FPCM more reliable for clustering complex datasets and better suited for analyzing TB-related factors across diverse regions.

One study implemented two fuzzy methods, FCM and FPCM, and found that FPCM performed better than FCM (20). Furthermore, a previous study also compared FCM and FPCM, showing the superiority of FPCM based on the number of iterations and computation time (21,22). This study combines panel data regression analysis with the FPCM method to identify factors that affect TB

This study integrates panel regression with FPCM to examine provincial TB patterns and identify region-specific clusters aligned with Indonesia's End TB 2030 roadmap. The resulting clusters correspond to operational strategies, such as targeted HIV/TB integration, community-based prevention, or surveillance maintenance. FPCM is expected to provide more robust clustering to noise and outliers and produce more stable and accurate clustering. Identifying these factors and clustering provinces based on them is expected to help create more effective and specific TB treatments tailored to the characteristics of each cluster. This study aims to contribute to TB control strategies by identifying tailored interventions for specific regions, aligning with the global goal of ending the TB epidemic by 2030.

METHOD

Study Design, Setting, and Population

An ecological time-series study design was selected to capture province-level trends and contextual determinants across Indonesia between 2020–2022. A time-series ecological study is especially useful for examining and understanding temporal patterns and relationships within historical data (23,24). It is well-suited for evaluating the time-related dynamics of TB cases in a specific population (25). The study utilized data from all 34 Indonesian provinces, ensuring comprehensive coverage for robust analysis.

Variables, Data Source, and Study Size

This study's dependent variable (Y) is the number of TB cases (person). The independent variables hypothesized to influence TB cases include the number of healthcare services (units) as X_1, number of HIV cases (person) as X_2, smoking prevalence among individuals aged $\geq$ 15 years (%) as X_3, number of registered food management places such as restaurants and water depots (units) as X_4, and number of registered public places and facilities such as schools and markets (units) as X_5. Data for these variables were sourced from secondary datasets provided by Statistics Indonesia and the Health Profile of Indonesia 2021-2023.

Statistical Analysis

This study employs statistical methods to derive generalizations from the findings (26,27). Data analysis was performed using R (version 4.4.2) for statistical computations and QGIS (version 3.28.1) for spatial mapping and visualization. The analysis consisted of three main stages: descriptive analysis, panel data regression, and FPCM. Descriptive analysis included measures of central tendency and dispersion to summarize variable distributions (28). In addition, descriptive statistics aim to enable comparisons and support further analysis (29).

The panel data regression analysis began with a multicollinearity check using Variance Inflation Factor (VIF) values. A VIF below 10 indicates no multicollinearity. The next step involved selecting parameters and the best regression model among the Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM). The selection process utilized the Chow Test to compare CEM and FEM, the Hausman Test to compare FEM and REM, and the Lagrange Multiplier (LM) Test for comparing CEM and REM, as appropriate. CEM is estimated by Ordinary Least Square (OLS), FEM is estimated by Least Squares Dummy Variable (LSDV), and REM is estimated by Generalized Least Squares (GLS).

After selecting the optimal model, assumption testing was performed. Heteroscedasticity was assessed using the Breusch-Pagan Test, and autocorrelation was checked with the Breusch-Godfrey Test. In cases where these assumptions were violated, robust standard errors were applied. Finally, the significance of the model was tested using the F-test for simultaneous effects and the t-test for partial effects. In contrast, the coefficient of determination (R^2) was used to evaluate the model's explanatory power.

For the FPCM analysis, data preprocessing involved calculating each province's average independent variable over three years. Average values from 2020–2022 were used in FPCM to represent stable provincial profiles and minimize short-term fluctuations, consistent with clustering practices in public health applications. As recommended, we conducted a sensitivity check using median values, which produced similar clustering patterns, indicating robustness of the chosen approach

The optimal cluster number was determined using Silhouette graphs, total within sum of squares, and cluster validity indices, including PEI, PCI, Kwon, and CL indices. Provinces were then clustered based on the optimal number of clusters, using predefined parameters such as a fuzzy weight of 2, a possibilistic weight of 2, a maximum of 1,000 iterations, an expected error threshold of 1×10^{-9}, and an initial objective function set to 0. The cluster characteristics were further analyzed using descriptive statistics, boxplot visualizations, and thematic maps to identify distinguishing features among the clusters. This comprehensive approach provided valuable insights into the factors associated TB and regional clustering patterns.

Ethical Approval

This study did not require ethical approval as it did not involve human participants and solely the analysis of publicly available data. This exemption is in accordance with the Indonesian National Guidelines on Health Research Ethics and institutional frameworks regarding the analysis of non-identifiable public domain datasets.

RESULTS

Descriptive Statistics Analysis

Table 1 presents the descriptive statistical analysis of variables by year, highlighting several concerning trends. From 2020 to 2022, Indonesia saw a significant rise in TB cases (Y), with averages increasing from 10,351 in 2020 to 11,688 in 2021 and 21,303 in 2022. The standard deviation also grew from 15,817 to 34,914, reflecting greater variability in TB case distribution across provinces, signaling challenges in TB management.

Table 1. Descriptive statistical analysis results on each variable per year

Year	Symbol	Minimum	Maximum	Mean	SD	IQR
2020	Y	918	79,423	10,351	15,817	6,852
	X_1	66	1,460	388	340	225
	X_2	27	7,157	1,235	1,824	887
	X_3	20.50	33.43	27.73	2.87	4.13
	X_4	632	22,081	5,359	4,632	4,398
	X_5	15	2,724	683	700	819
2021	Y	995	91,368	11,688	17,675	8,307
	X_1	68	1,474	392	342	229
	X_2	31	5,872	1,085	1,556	981
	X_3	19.58	34.07	28.00	3.06	3.43
	X_4	712	24,217	6,025	5,311	4,441
	X_5	26	13,723	2,319	2,857	1,850
2022	Y	1,738	184,406	21,303	34,914	13,040
	X_1	97	1,317	395	290	188
	X_2	38	8,680	1,558	2,147	1,012
	X_3	17.91	33.81	26.76	3.73	4.95
	X_4	1,321	45,986	10,003	9,280	6,742
	X_5	128	6,708	1,243	1,514	842

Where: Y = number of TB cases, X_1 = number of healthcare services, X_2 = number of HIV cases, X_3 = smoking prevalence among individuals aged ≥ 15 years, X_4 = number of registered food management places, and X_5 = number of registered public places and facilities

Healthcare services (X_1) increased marginally from 388 units in 2020 to 395 units in 2022, showing limited progress in addressing TB challenges. Despite this, the rise in health services was insufficient to address the surge in TB cases. HIV cases (X_2) fluctuated, dropping from 1,235 in 2020 to 1,085 in 2021 before rising to 1,558 in 2022. Smoking prevalence (X_3) increased from 27.73% in 2020 to 28.00% in 2021, followed by a slight decline to 26.76% in 2022. These fluctuations suggest a need for more comprehensive public health strategies.

The number of food management premises (X_4) and public facilities (X_5) showed significant growth. Registered food management premises nearly doubled from 5,359 in 2020 to 10,003 in 2022, reflecting economic growth but also potential public health concerns. Public facilities rose from 683 in 2020 to 2,319 in 2021, decreasing to 1,243 in 2022. This growth reflects economic development and urbanization but poses challenges to maintaining hygiene and controlling TB transmission.

Panel Data Regression

Multicollinearity Check

The multicollinearity check showed VIF values of 2.68 for healthcare services (X_1), 8.71 for HIV cases (X_2), 1.44 for smoking prevalence (X_3), 5.35 for food management places (X_4), and 2.08 for public places and facilities (X_5). Since all VIF values are below 10, there is no multicollinearity among the independent variables, indicating that the predictors are not excessively correlated.

Model Selection

Parameter estimations for these approaches are represented by equation 1 (CEM), equation 2 (FEM), and equation 3 (REM). The panel data regression model selection involved several tests. The Chow test favored FEM over CEM (p = 0.000), and the Hausman test confirmed FEM as superior to REM (p = 0.000). Based on the model selection analysis, FEM is utilized for further analysis in this study. FEM outperformed CEM and REM by accounting for unobserved provincial heterogeneity that is time-invariant, consistent with panel data theory

$$\hat{y}_{it} = -47.397 - 7.1X_{1it} + 6.22X_{2it} - 1.549,5X_{3it} + 2.11X_{4it} - 0.79X_{5it} \quad (1)$$
$$\hat{y}_{it} = \beta_{0i} - 8.374X_{1it} + 13.75X_{2it} + 905.32X_{3it} + 1.62X_{4it} + 1.11X_{5it} \quad (2)$$
$$\hat{y}_{it} = -36{,}298 - 12.85X_{1it} + 7.23X_{2it} + 1{,}142.8X_{3it} + 2.05X_{4it} + 0.27X_{5it} \quad (3)$$

Assumption Testing

The Breusch-Pagan test identified heteroscedasticity (p = 0.000), and the Breusch-Godfrey test detected autocorrelation (p = 0.000). Robust standard errors were applied to address these issues. A p-value of 0.000 ($p < 0.05$) from the result of the simultaneous test indicates that at least one independent variable significantly influences TB cases in Indonesia. Further, based on the partial test shows that all independent variables analyzed had a significant effect on the dependent variable (Y). The corrected estimation results with robust standard errors and the simultaneous test and partial test analysis results can be seen in Table 2.

Tabel 2. Result of Assumption Testing

Variable	Estimation	*Standard error*	*F-test*	*t-value*	*p-value*
X_1	-8.37	1.71		-4.90	0.00
X_2	13.76	3.24		4.24	0.00
X_3	905.32	423.80	0.05	2.14	0.04
X_4	1.62	0.30		5.42	0.00
X_5	1.11	0.47		2.38	0.02

In the FEM, the intercept coefficient (β_0) and the dummy coefficients (β_i). Vary across provinces. The purpose of Entity Dummy Variables in the FEM for panel data regression is to control for unobserved heterogeneity that is time-invariant across entities. These variables ensure that the model accounts for individual-specific characteristics that do not change over time but might influence the dependent variable. The intercept coefficient (β_0) and dummy coefficient (β_i) for each province with Aceh Province as a reference category can be seen in Table 3. So, the final model for the panel data regression method using FEM is as follows

$$\hat{y}_{it} = \beta_{01} + \beta_i - 8.37X_{1it} + 13.76X_{2it} + 905.32X_{3it} + 1.62X_{4it} + 1.11X_{5it} \quad (3)$$

Table 3. Intercept (β_{0i}) and entity dummy coefficients (β_{1i}) in FEM for each province

No.	Province	$\boldsymbol{\beta_{0i}}$	$\boldsymbol{\beta_{1i}}$	No.	Province	$\boldsymbol{\beta_{0i}}$	$\boldsymbol{\beta_{1i}}$
1	Aceh	- 26.398	-	18	West Nusa Tenggara	-33.882	-7.484
2	North Sumatra	-32.920	-6.522	19	East Nusa Tenggara	-30.035	-3.636
3	West Sumatra	-37.175	-10.777	20	West Kalimantan	-33.862	-74.64

No.	Province	β_{0i}	β_{1i}	No.	Province	β_{0i}	β_{1i}
4	Riau	-36.910	-10.512	21	Central Kalimantan	-32.237	-5.838
5	Jambi	-31.208	-4.810	22	South Kalimantan	-35.309	-8.911
6	South Sumatra	-25.317	1.080	23	East Kalimantan	-48.172	-21.774
7	Bengkulu	-34.262	-7.864	24	North Kalimantan	-28.364	-19.66
8	Lampung	-35.746	-9.348	25	North Sulawesi	-27.531	-11.33
9	Bangka Belitung	-30.887	-4.489	26	Central Sulawesi	-33.636	-7.238
10	Riau Islands	-37.214	-10.816	27	South Sulawesi	-43.956	-17.558
11	Jakarta	-75.552	-49.154	28	Southeast Sulawesi	-29.649	-3.251
12	West Java	-47.135	-69.432	29	Gorontalo	-28.474	-2.076
13	Central Java	-75.205	-48.806	30	West Sulawesi	-24.091	2.306
14	Yogyakarta	-29.631	-3.233	31	Maluku	-27.586	-1.188
15	East Java	-95.830	-20.737	32	North Maluku	-29.051	-2.653
16	Banten	-35.552	-9.154	33	West Papua	-29.159	-2.761
17	Bali	-43.796	-10.816	34	Papua	-48.825	-22.427

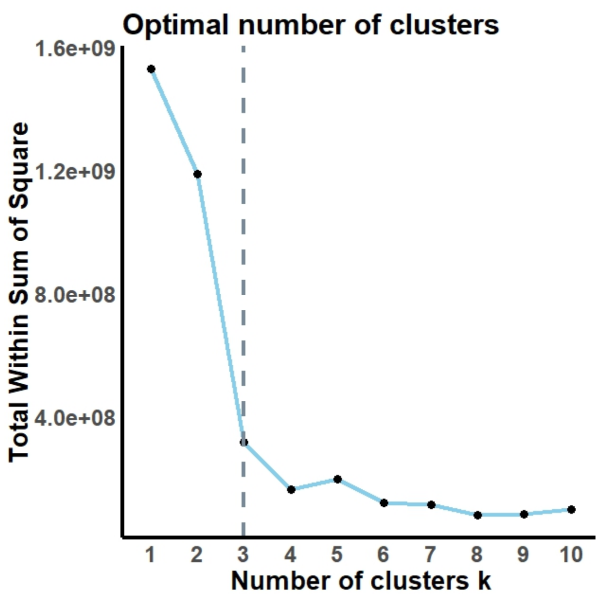
For healthcare services (X_1), a coefficient of -8.37 indicates that an increase of one unit in the number of health services reduces TB cases by approximately eight people. Conversely, the coefficient for HIV cases (X_2) is 13.76, suggesting that each additional HIV case increases TB cases by around 14 people. Similarly, smoking prevalence (X_3) has a substantial impact, with a coefficient of 905.32, indicating that a 1% increase in smoking prevalence raises TB cases by approximately 905 people. The number of food management places (X_4) has a coefficient of 1.62, meaning that each additional unit increases TB cases by about two people. Lastly, public places and facilities (X_5) have a coefficient of 1.11, signifying that an increase of one-unit results in approximately one additional TB case.

Based on the FEM model, the R^2 of 0.926 and an adjusted R^2 of 0.88 2 were obtained. This R^2 indicates that the independent variables in the model can explain 92.6% of the variability in the number of TB cases. In contrast, the adjusted R^2 of 0.882 indicates the R^2 adjustment for the number of variables in the model. Overall, the high R^2 and adjusted R^2 values indicate that the panel regression model perfectly fits the data and can explain most of the variability in the number of TB cases. It indicates that the independent variables selected in the model are highly relevant and contribute significantly to predicting the number of TB cases.

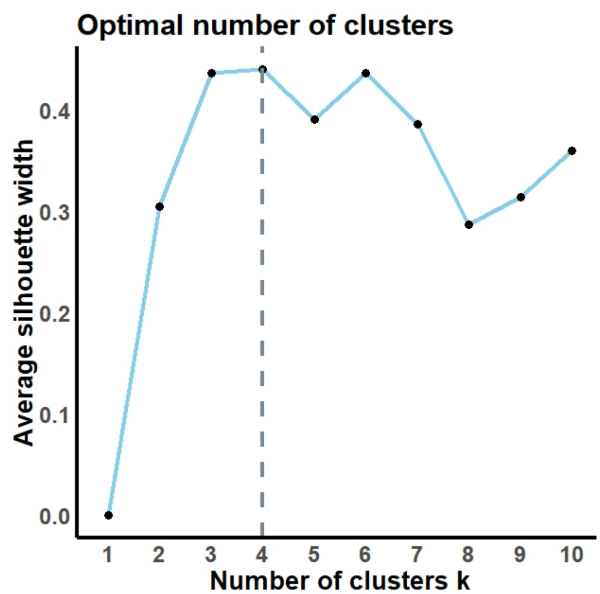
Fuzzy Possibilistic C-Means Clustering (FPCM)

Cluster Optimization

The optimal number of clusters was determined using the Total Within Sum Square (TWSS), Silhouette method, and cluster validity indices. TWSS suggested three clusters (Figure 1 (a)), while the Silhouette method favored four (Figure 1(b)). Validity indices (PEI, Kwon, PCI, and CL) were analyzed to resolve this. For three clusters: PEI at 0.35, Kwon at 7.87, PCI at 0.81, CL at 0.77; for four clusters: PEI at 0.39, Kwon at 8.47, PCI at 0.79, CL at 0.74. Three clusters were chosen as optimal based on the highest PCI and Kwon values and the lowest PEI and CL.



(a)



(b)

Figure 1. (a) TWSS graph and (b) Silhouette graph to determine the optimum number of clusters

Clustering process

Fuzzy clustering using the FPCM method grouped Indonesia's provinces into three clusters based on their possibilistic membership degrees (Table 4). Provinces with the highest membership in a cluster share its defining characteristics, such as Aceh in Cluster 1. Cluster 1 includes 12 provinces (e.g., Aceh, North Sumatra), Cluster 2 contains 17 provinces (e.g., Jambi, South Sumatra), and Cluster 3 comprises five provinces (e.g., Jakarta, West Java). Thematic maps of these clusters are shown in Figure 2. Thematic maps in Figure 2 illustrate the spatial distribution of TB-related factors, aiding in visually identifying high-priority regions.

Table 4. Possibilistic membership degrees

No.	Province	Cluster 1	Cluster 2	Cluster 3	Final Cluster
1	Aceh	0.016295	0.005609	0.005472	1
2	North Sumatra	0.021810	0.002929	0.006232	1
3	West Sumatra	0.267201	0.001300	0.007565	1
4	Riau	0.158702	0.001557	0.007262	1
5	Jambi	0.010328	0.014188	0.005050	2
6	South Sumatra	0.006360	0.232435	0.004553	2
7	Bengkulu	0.005234	0.092882	0.004318	2
8	Lampung	0.082606	0.002035	0.006724	1
9	Bangka Belitung Islands	0.004808	0.098884	0.004187	2
10	Riau Islands	0.014020	0.007492	0.005363	1
11	Jakarta	0.005046	0.000298	0.017657	3
12	West Java	0.000356	0.000047	0.033096	3
13	Central Java	0.001281	0.000123	0.132124	3
14	Yogyakarta	0.006015	0.230269	0.004500	2
15	East Java	0.000826	0.000091	0.649886	3

No.	Province	Cluster 1	Cluster 2	Cluster 3	Final Cluster
16	Banten	0.058233	0.001979	0.006953	1
17	Bali	0.015057	0.005705	0.005600	1
18	West Nusa Tenggara	0.008403	0.031383	0.004840	2
19	East Nusa Tenggara	0.006324	0.045742	0.004590	2
20	West Kalimantan	0.015127	0.007478	0.005438	1
21	Central Kalimantan	0.008090	0.036846	0.004779	2
22	South Kalimantan	0.192674	0.001218	0.007553	1
23	East Kalimantan	0.038072	0.000610	0.010324	1
24	North Kalimantan	0.004477	0.054542	0.004099	2
25	North Sulawesi	0.003829	0.022128	0.003937	2
26	Central Sulawesi	0.008440	0.030647	0.004837	2
27	South Sulawesi	0.008803	0.000357	0.015202	3
28	Southeast Sulawesi	0.011674	0.011248	0.005176	1
29	Gorontalo	0.003795	0.022893	0.003922	2
30	West Sulawesi	0.003143	0.009733	0.003700	2
31	Maluku	0.002892	0.007270	0.003608	2
32	North Maluku	0.002514	0.004615	0.003436	2
33	West Papua	0.002863	0.006971	0.003603	2
34	Papua	0.004701	0.008497	0.004415	2

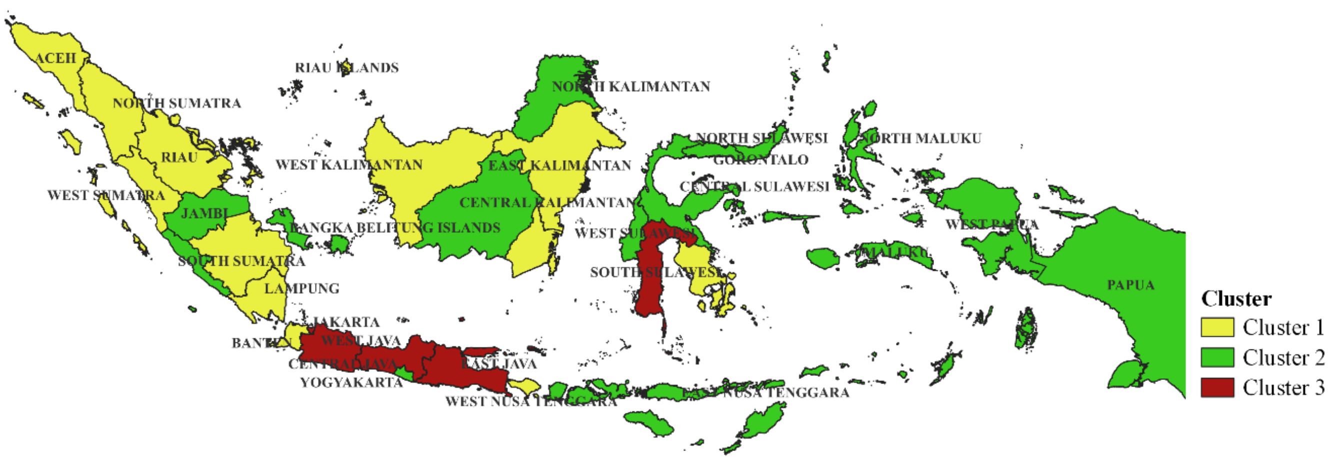


Figure 2. Thematic map of provinces in Indonesia based on FPCM

Clustering Analysis

The main focus on each variable is based on its influence on the increase in TB cases from the previously formed panel data regression model. If a variable positively influences the increase in TB cases, then the cluster focused on that variable is the cluster with the highest interquartile range and median values. This reason aims to determine the appropriate treatment in each cluster to reduce the number of TB cases. Figure 3(a) illustrates that Cluster 3 has the widest interquartile range and highest median for the number of health services, followed by Cluster 1, with Cluster 2 showing the lowest values. Improving health services in Cluster 2 should be prioritized to help eliminate TB cases while equalizing the distribution of health services in Cluster 1 will reduce TB cases.

Figure 3(b) reveals that HIV cases are most prevalent in Cluster 3, with a high interquartile range and median, followed by Clusters 1 and 2. It underscores the need for special improved HIV treatment in Cluster 3 to mitigate TB cases. Figure 3(c) shows minimal variation in smoking rates among clusters, but Cluster 2 has the highest interquartile range and median. Reducing smoking rates in Cluster 2 should be prioritized.

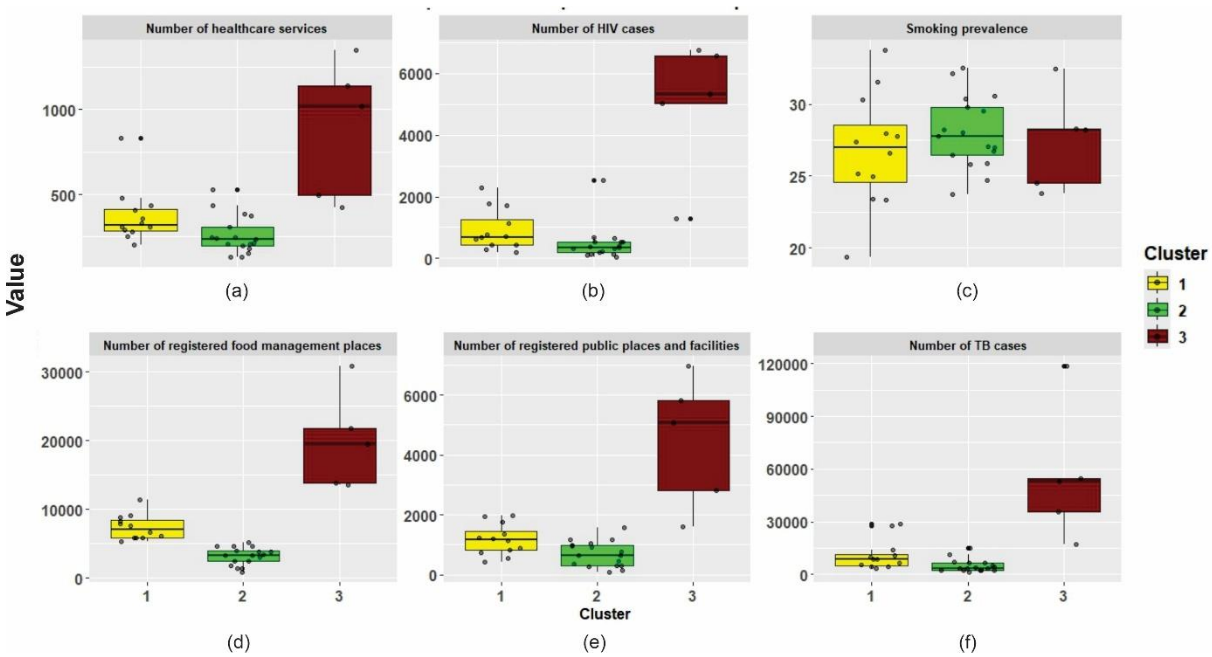


Figure 3. Distribution of Variables by Cluster using boxplot

Figures 3(d) and 3(e) indicate that Cluster 3 has the highest interquartile range and median for food management places and public facilities. Since public places can promote TB transmission through increased interactions, improving health standards in these locations, particularly Cluster 3, is essential. Furthermore, based on the distribution of TB cases in each cluster shown in Figure 3(f). Cluster 3 is the top priority in reducing TB cases in Indonesia. This result is evident from the wide interquartile range and high median. This finding is in line with the distribution of the variables in the panel data regression model as equation 3, where in addition to the importance of equitable distribution of health services, special attention is needed to the number of HIV cases, smoking prevalence, the number of registered food management places, and the number of registered public places and facilities.

DISCUSSION

An increase in healthcare services has shown significant potential to enhance TB management through improved early diagnosis and treatment accessibility, with more health services reducing the number of TB cases. This finding is consistent with several studies where Increasing health services helps reduce TB cases by improving access to early detection, diagnosis, and treatment (30–32). More services encourage individuals to seek care, receive timely treatment, and reduce transmission risks. Additionally, health services offer better education and preventive measures, such as vaccination and treatment adherence, which further control TB spread. Enhanced health infrastructure also aids in managing co-morbidities that can complicate TB treatment and increase susceptibility to infection (33,34).

Conversely, higher HIV prevalence exacerbates TB incidence by compromising immune responses and complicating treatment efficacy, where an increase in HIV cases tends to increase the number of TB cases. This finding is in line with the previous studies, which found a significant correlation between HIV and the number of TB cases (35,36). Similarly, Lee et al. (2021) noted that HIV infection makes TB difficult to find, more difficult to treat, or more resistant to certain drugs (37). In addition, smoking prevalence significantly exacerbates TB risk by weakening lung function and increasing susceptibility to infection and number of TB cases. This finding is consistent with the study by Feng et al. (2023) and Quan et al. (2022), who concluded that TB incidence can be reduced through smoking cessation (38,39), as well as several studies that stated that smoking can increase the incidence of TB and more susceptible to TB infection (40,41). Based on finding of previous studies, this necessitates situating findings within the strategic objectives of Indonesia's Health Strategic Plan 2025–2029 and WHO's four pillars of TB control, emphasizing integrated care and supportive systems.

This study proves that TB is not only influenced by health factors but also by non-health factors, such as the number of food management places and public facilities that negatively impact TB. The relationship between food management sites, public facilities, and the incidence of TB is influenced by various socioeconomic and environmental factors. Increased food management places and public places can contribute to higher TB cases due to factors such as overcrowding, poor ventilation, and inadequate health resources, which facilitate the transmission of Mycobacterium tuberculosis. Moreover, environmental factors linked to food management and public facilities can also influence TB transmission.

Some studies discussed how environmental health factors, including environmental exposure, inadequate housing, and high population density, have been repeatedly linked to TB transmission, particularly in settings with poor ventilation and overcrowding, and are critical to the incidence of TB (42,43). Poorly managed public spaces can lead to overcrowding and inadequate air quality, known risk factors for TB transmission (44,45). Therefore, improved health standards in public places are needed to reduce the potential spread of diseases, particularly TB.

Identifying provinces that require priority treatment based on factors affecting TB is important to assist the government in formulating appropriate and effective policies. Identifying provinces that require priority treatment based on factors affecting TB helps the government allocate resources efficiently and implement targeted interventions where they are most needed. It ensures that policies address each region's specific health, socioeconomic, and environmental challenges, maximizing their impact in controlling TB (46,47).

Based on clustering analysis using FPCM, provinces in Indonesia can be divided into three clusters with different characteristics. Furthermore, these membership degrees of FPCM indicate prioritization gradients for policy attention, where higher membership in three Cluster denotes an urgent need for cross-sectoral interventions. This finding helps identify provinces that require priority treatment. Cluster 3 stands out in almost all variables, especially the number of HIV cases, so addressing HIV cases needs to be prioritized in Cluster 3. In addition, the higher number of food management sites and public facilities in Cluster 3 requires improved health standards in public places, especially in Cluster 3. This finding supports the association that improved access to health services is linked to lower TB incidence. Meanwhile, Cluster 2 requires improved health services and measures to reduce the percentage of smokers. Cluster 1 falls between Clusters 2 and 3 in all variables. The cluster priorities are Cluster 3, Cluster 1, and Cluster 2.

Overall, these findings emphasize the importance of improving the quality of the health sector, including access to health facilities, addressing diseases such as HIV, controlling smoking, and improving health standards in food management and public facilities to tackle TB in Indonesia. These health sectors were positively associated with TB incidence, consistent with prior evidence (48–50). These measures should be tailored to the specific needs of each cluster. However, this study has several limitations that need to be considered. The limited period and variables may limit a thorough understanding of the factors influencing TB. Therefore, future research should extend the period and consider adding variables to identify other factors influencing TB. Futhermore, the ecological fallacy must be explicitly acknowledged; these ecological models cannot infer individual-level causality, reinforcing that conclusions are limited to regional determinants.

CONCLUSION

This study underscores the multifaceted factors influencing TB incidence in Indonesia, highlighting the critical roles of health services, HIV cases, smoking prevalence, food management places, and public facilities. The findings illustrate that increasing the number of healthcare services can significantly reduce TB cases by improving early detection and treatment access. At the same time, HIV and smoking prevalence exacerbate the disease's spread. Additionally, the growing number of food management sites and public facilities poses environmental risks that necessitate stricter health standards to control TB transmission.

Using FPCM, provinces were grouped into three distinct clusters, each requiring tailored interventions. Cluster 3 emerged as the most critical region requiring intervention due to its high HIV prevalence and density of public facilities contributing to TB transmission. Cluster 2 requires enhanced healthcare services and smoking cessation efforts, while Cluster 1 presents moderate challenges across variables. These insights provide a framework for targeted policymaking, enabling resource allocation that reflects each province's unique characteristics. The findings conclude with a precise policy recommendation promoting a cluster-based allocation mechanism and systematic monitoring of provincial TB indicators.

AUTHOR'S CONTRIBUTION STATEMENT

NRS: Concepts, Design, Definition of intellectual content, Data analysis, Statistical analysis, Manuscript preparation, Manuscript editing, Manuscript review; MS, MK: Concepts, Design Data analysis, Statistical analysis, Manuscript preparation; LR, ZMK, SM, RK: Manuscript preparation and editing, Manuscript review; VC: Manuscript preparation, Manuscript editing, Manuscript review; MINA: Manuscript editing, Manuscript review.

CONFLICTS OF INTEREST

The authors have no conflicts of interest.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

This manuscript was developed without the use of Generative AI or AI-assisted technologies at any stage. The writing, idea generation, image production, graphical elements, data collection, and analysis were all conducted manually.

SOURCE OF FUNDING STATEMENTS

Research reported in this publication was supported by the Fogarty International Center and the National Institute of Allergy and Infectious Diseases, of the National Institutes of Health under Award Number D43 TW009522. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ACKNOWLEDGMENTS

Thank you to the Statistics Indonesia and the Health Profile of Indonesia for their support and provision of data, which was crucial for this study.

BIBLIOGRAPHY

1. Antabe R, Ziegler BR. Diseases, Emerging and Infectious. In: International Encyclopedia of Human Geography. 2nd ed. 2020;3:389–391. https://doi.org/10.1016/B978-0-08-102295-5.10439-1
2. Saputra A, Sofyan H, Kesuma ZM, Sasmita NR, Wichaidit W, Chongsuvivatwong V. Spatial Patterns of Tuberculosis in Aceh Province During the COVID-19 Pandemic: A Geospatial Autocorrelation Assessment. IOP Conference Series: Earth and Environmental Science. 2024;1356:012099. https://doi.org/10.1088/1755-1315/1356/1/012099
3. World Health Organization. Global Tuberculosis Report 2024. Geneva: WHO; 2024.
4. Sasmita NR, Arifin M, Kesuma ZM, Rahayu L, Mardalena S, Kruba R. Spatial Estimation for Tuberculosis Relative Risk in Aceh Province, Indonesia: A Bayesian Conditional Autoregressive Approach with the Besag-York-Mollie (BYM) Model. Journal of Applied Data Science. 2024;5:342–356. https://doi.org/10.47738/jads.v5i2.185
5. Sasmita NR, Ramadani M, Ikhwan M, Munawwarah M, Rahayu L, Mardalena S, Asshiddiqi MIN, Suyanto S, Safira N. A Stochastic Projection for Tuberculosis Elimination in Indonesia by 2030. Media Publikasi Promosi Kesehatan Indonesia. 2025;8:1297–1314. https://doi.org/10.56338/mppki.v8i11.8548
6. Ministry of Health Indonesia. Indonesia Health Profile 2022. Jakarta: Ministry of Health Indonesia; 2022.

7. Sasmita NR, Phonna RA, Kesuma ZM, Kamal S, Yusya N. Spatial-Temporal Epidemiology of COVID-19 in Aceh, Indonesia: A Statistical Perspective. Unnes Journal of Public Health. 2024;13:67–79. https://doi.org/10.15294/ujph.v13i2.3428
8. Rojali R, Sari DI. Relationship of Individual Characteristics, Physical Home Environment and Behavior with the Incidence of Pulmonary TB in Cijoro Pasir Village, Muara Village East Ciujung and West Rangkasbitung Village, Rangkasbitung Subdistrict, Lebak Regency 2019. ADI Journal on Recent Innovation. 2020;1:167–179. https://doi.org/10.34306/ajri.v1i2.36
9. Fahdhienie F, Mudatsir M, Abidin TF, Nurjannah N. Risk Factors of Pulmonary Tuberculosis in Indonesia: A Case-Control Study in a High Disease Prevalence Region. Narra Journal. 2024;4:1–12. https://doi.org/10.52225/narra.v4i2.943
10. Liyew AM, Clements ACA, Akalu TY, Gilmour B, Alene KA. Ecological-Level Factors Associated with Tuberculosis Incidence and Mortality: A Systematic Review and Meta-Analysis. PLOS Global Public Health. 2024;4:1–22. https://doi.org/10.1371/journal.pgph.0003425
11. Gupta P, Singh P, Das A, Kumar R. Determinants of Tuberculosis: An Example of High Tuberculosis Burden in the Saharia Tribe. Frontiers in Public Health. 2023;11:1226980. https://doi.org/10.3389/fpubh.2023.1226980
12. Lima LV de, Pavinati G, Palmieri IGS, Bernardo PHP, Santos VMA dos, Gomes MF, et al. An Ecological Study Mapping Socioeconomic Inequalities in Tuberculosis Incidence in a Southern State of Brazil. Spatial and Spatio-Temporal Epidemiology. 2024;51:100689. https://doi.org/10.1016/j.sste.2024.100689
13. Ningrum DP, Saifudin T, Suliyanto S, Chamidah N. Tuberculosis Modeling in East Java Based on Geographically Weighted Regression Approach. Jurnal Matematika, Statistika dan Komputasi. 2022;19:19–32. https://doi.org/10.20956/j.v19i1.21262
14. Wooldridge JM. Introductory Econometrics: A Modern Approach. 6th ed. Canada: Cengage Learning; 2016.
15. Gui S, Zhang X, Sun Z, Yao Y. Uncovering Spatiotemporal Development Patterns of AIDS in China: A Study Using Panel Data with Joinpoint Regression Analysis and Spatial Clustering. Health & Place. 2024;90:103353. https://doi.org/10.1016/j.healthplace.2024.103353
16. Thakur PS, Verma RK, Tiwari R. A Comparison Between the Fuzzy C-Means Clustering Algorithm and the K-Mean Clustering Algorithm. Engineering Mathematics Letters. 2024;1:1–11. https://doi.org/10.28919/eml/8403
17. Li H, Wang J. From Soft Clustering to Hard Clustering: A Collaborative Annealing Fuzzy C-Means Algorithm. IEEE Transactions on Fuzzy Systems. 2024;32:1181–1194. https://doi.org/10.1109/TFUZZ.2023.3319663
18. Pushkarova Y, Kholodniuk P. A New Procedure for Unsupervised Clustering Based on Combination of Artificial Neural Networks. European Journal of Artificial Intelligence and Machine Learning. 2023;2:1–3. https://doi.org/10.24018/ejai.2023.2.4.31
19. Pratiwi NBI, Saputro DRS. Fuzzy C-Shells Clustering Algorithm. Journal of Physics: Conference Series. 2020;1613:012006. https://doi.org/10.1088/1742-6596/1613/1/012006
20. Kurniasari D, Kurniawati V, Nuryaman A, Usman M, Nisa RK. Implementation of Fuzzy C-Means and Fuzzy Possibilistic C-Means Algorithms on Poverty Data in Indonesia. BAREKENG: Jurnal Ilmu Matematika dan Terapan. 2024;18:1919–1930. https://doi.org/10.30598/barekengvol18iss3pp1919-1930
21. Putri GNS, Ispriyanti D, Widiharih T. Implementasi Algoritma Fuzzy C-Means dan Fuzzy Possibilistic C-Means untuk Klasterisasi Data Tweets pada Akun Twitter Tokopedia. Jurnal Gaussian. 2022;11:86–98. https://doi.org/10.14710/j.gauss.v11i1.33996
22. Özdemir O, Kaya A. Comparison of FCM, PCM, FPCM and PFCM Algorithms in Clustering Methods. Afyon Kocatepe University Journal of Science and Engineering. 2019;19:92–102. https://doi.org/10.35414/akufemubid.429540
23. Carlos de Abreu L, Elmusharaf K, Eduardo Gomes Siqueira C. A Time-Series Ecological Study Protocol to Analyze Trends of Incidence, Mortality, Lethality of COVID-19 in Brazil. Journal of Human Growth and Development. 2021;31:491–495. https://doi.org/10.36311/jhgd.v31.12667

24. Walia GK, Mandal S, Jaganathan S, Jaacks LM, Sieber NL, Dhillon PK, et al. Leveraging Existing Cohorts to Study Health Effects of Air Pollution on Cardiometabolic Disorders: India Global Environmental and Occupational Health Hub. Environmental Health Insights. 2020;14:1–6. https://doi.org/10.1177/1178630220915688
25. Sasmita NR, Khairul M, Fikri MK, Rahayu L, Kesuma ZM, Mardalena S, et al. Relative Risk and Distribution Assessment of Tuberculosis Cases: A Time-Series Ecological Study in Aceh, Indonesia. Media Publikasi Promosi Kesehatan Indonesia. 2025;8:407–417. https://doi.org/10.56338/mppki.v8i6.7264
26. Azharuddin, Sasmita NR, Idroes GM, Andid R, Raihan, Fadlilah T, et al. Patient Satisfaction and Its Socio-Demographic Correlates in Zainoel Abidin Hospital, Indonesia: A Cross-Sectional Study. Unnes Journal of Public Health. 2023;12:57–67. https://doi.org/10.15294/ujph.v12i2.69233
27. Sasaki D, Sofyan H, Sasmita NR, Affan M, Nizamuddin N. Assessing the Intermediate Function of Local Academic Institutions During the Rehabilitation and Reconstruction of Aceh, Indonesia. Journal of Disaster Research. 2021;16:1265–1273. https://doi.org/10.20965/jdr.2021.p1265
28. Earlia N, Bulqiah M, Muslem M, Karma T, Suhendra R, Maulana A, et al. Protective Effects of Acehnese Traditionally Fermented Coconut Oil (Pliek U Oil) and Its Residue (Pliek U) in Ointment Against UV Light Exposure: Studies on Male Wistar Rat Skin (Rattus norvegicus). Sains Malaysiana. 2021;50:1285–1295. https://doi.org/10.17576/jsm-2021-5005-09
29. Rahayu L, Sasmita NR, Adila WF, Kesuma ZM, Kruba R. Spatial Estimation of Relative Risk for Dengue Fever in Aceh Province Using Conditional Autoregressive Method. Journal of Applied Data Science. 2023;4:466–479. https://doi.org/10.47738/jads.v4i4.141
30. Jesus GS, Pescarini JM, Silva AF, Torrens A, Carvalho WM, Junior EPP, et al. The Effect of Primary Health Care on Tuberculosis in a Nationwide Cohort of 7.3 Million Brazilian People: A Quasi-Experimental Study. The Lancet Global Health. 2022;10:e390–e397. https://doi.org/10.1016/S2214-109X(21)00550-7
31. Villalva-Serra K, Barreto-Duarte B, Rodrigues MM, Queiroz ATL, Martinez L, Croda J, et al. Impact of Strategic Public Health Interventions to Reduce Tuberculosis Incidence in Brazil: A Bayesian Structural Time-Series Scenario Analysis. The Lancet Regional Health – Americas. 2025;41:100963. https://doi.org/10.1016/j.lana.2024.100963
32. Deng L, Zhao F, Li Z, Zhang W, He G, Ren X. Epidemiological Characteristics of Tuberculosis Incidence and Its Macro-Influence Factors in Chinese Mainland During 2014–2021. Infectious Diseases of Poverty. 2024;13:1–15. https://doi.org/10.1186/s40249-024-01203-6
33. Wulandari LPL, Negara SNS, Wahyuningtias SD, Mashuri YA, Putra IWCSD, Fardousi N, et al. Delivering HIV and TB Services Amidst the COVID-19 Pandemic in Indonesia: A Qualitative Study of Challenges and Mitigation Strategies. Journal of Global Health Reports. 2024;8:1–14. https://doi.org/10.29392/001c.117620
34. Dlatu N, Longo-Mbenza B, Apalata T. Models of Integration of TB and HIV Services and Factors Associated with Perceived Quality of TB-HIV Integrated Service Delivery in O.R. Tambo District, South Africa. BMC Health Services Research. 2023;23:1–14. https://doi.org/10.1186/s12913-023-09748-2
35. Song Y, Jin Q, Qiu J, Ye D. A Systematic Review and Meta-Analysis on the Correlation Between HIV Infection and Multidrug-Resistance Tuberculosis. Heliyon. 2023;9:e21956. https://doi.org/10.1016/j.heliyon.2023.e21956
36. Lelisho ME, Wotale TW, Tareke SA, Alemu BD, Hassen SS, Yemane DM, et al. Survival Rate and Predictors of Mortality Among TB/HIV Co-Infected Adult Patients: Retrospective Cohort Study. Scientific Reports. 2022;12:1–11. https://doi.org/10.1038/s41598-022-23316-4
37. Lee CM, Lee E, Bang JH, Park S, Park WB, Oh M, et al. Prevalence of Multidrug-Resistant Tuberculosis in HIV/Tuberculosis Co-Infected Patients. Infection & Chemotherapy. 2021;53:792–795. https://doi.org/10.3947/ic.2021.0085
38. Feng Y, Xu Y, Yang Y, Yi G, Su H, Chen H, et al. Effects of Smoking on the Severity and Transmission of Pulmonary Tuberculosis: A Hospital-Based Case Control Study. Frontiers in Public Health. 2023;11:1017967. https://doi.org/10.3389/fpubh.2023.1017967
39. Quan DH, Kwong AJ, Hansbro PM, Britton WJ. No Smoke Without Fire: The Impact of Cigarette Smoking on the Immune Control of Tuberculosis. European Respiratory Review. 2022;31:210252. https://doi.org/10.1183/16000617.0252-2021

40. de Vargas KR, Freitas AA, Azeredo ACV, Silva DR. Smoking Prevalence and Effects on Treatment Outcomes in Patients with Tuberculosis. Revista da Associação Médica Brasileira. 2021;67:406–410. https://doi.org/10.1590/1806-9282.20200825
41. Bay JG, Patsche CB, Svendsen NM, Gomes VF, Rudolf F, Wejse C. Tobacco Smoking Impact on Tuberculosis Treatment Outcome: An Observational Study from West Africa. International Journal of Infectious Diseases. 2022;124:50–55. https://doi.org/10.1016/j.ijid.2022.07.067
42. Sulidah S, Irwan M, Elmania E, Fadlilah S, Hamdani Rahil N, Nugroho A. Home Environment as a Risk Factor for Increased Incidence of Tuberculosis: A Case-Control Study. Ciencia y Enfermería. 2024;30:1–9. https://doi.org/10.29393/CE30-5HESA60005
43. Tanjung R, Mahyuni EL, Tanjung N, Simarmata OS, Sinaga J, Nolia HR. The Spatial Distribution of Pulmonary Tuberculosis in Kabanjahe District, Karo Regency, Indonesia. Open Access Macedonian Journal of Medical Sciences. 2021;9:817–822. https://doi.org/10.3889/oamjms.2021.6808
44. Lee JY, Kwon N, Goo G, Cho S. Inadequate Housing and Pulmonary Tuberculosis: A Systematic Review. BMC Public Health. 2022;22:1–22. https://doi.org/10.1186/s12889-022-12879-6
45. Paleckyte A, Dissanayake O, Mpagama S, Lipman MC, McHugh TD. Reducing the Risk of Tuberculosis Transmission for Healthcare Workers in High-Incidence Settings. Antimicrobial Resistance & Infection Control. 2021;10:1–11. https://doi.org/10.1186/s13756-021-00975-y
46. Cole B, Nilsen DM, Will L, Etkind SC, Burgos M, Chorba T. Essential Components of a Public Health Tuberculosis Prevention, Control, and Elimination Program: Recommendations of the Advisory Council for the Elimination of Tuberculosis and the National Tuberculosis Controllers Association. MMWR Recommendations and Reports. 2020;69:1–27. https://doi.org/10.15585/mmwr.rr6907a1
47. van der Westhuizen HM, Giddy J, Coetzee R, Makanda G, Tisile P, Galloway M, et al. Strengthening Accountability for Tuberculosis Policy Implementation in South Africa: Perspectives from Policymakers, Civil Society, and Communities. BMC Global and Public Health. 2024;48:1–10. https://doi.org/10.1186/s44263-024-00077-y
48. Teshima A, Shatnawi AA, Satyanarayana S, Khader YS, Maia IF, Wilson NC. High Prevalence of Current Tobacco Smoking Among Patients with Tuberculosis and People Living with HIV in Jordan: A Cross-Sectional Survey. Tobacco Induced Diseases. 2023;21:1–13. https://doi.org/10.18332/tid/171551
49. Ojo T, Ruan C, Hameed T, Malburg C, Thunga S, Smith J, et al. HIV, Tuberculosis, and Food Insecurity in Africa: A Syndemics-Based Scoping Review. International Journal of Environmental Research and Public Health. 2022;19:1101. https://doi.org/10.3390/ijerph19031101
50. Arja A, Bogale B, Gebremedhin M. Health System Delay and Its Associated Factors Among Tuberculosis Patients in Gamo Zone Public Health Facilities, Southern Ethiopia: An Institution-Based Cross-Sectional Study. Journal of Clinical Tuberculosis and Other Mycobacterial Diseases. 2022;28:100325. https://doi.org/10.1016/j.jctube.2022.100325