Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 4 No. 1 (January) (2026)

Trend Analysis and Forecasting of Tuberculosis Cases in Pakistan Using Regression Analysis with Real clinical Data (2002–2018)

Dr. Zohaib Ali
Lecture, Institute of Mathematics & Computer Science**,** University of Sindh, Jamshoro, Pakistan Email: zohaib.ali@usindh.edu.pk

Dr. Syeda Hira Fatima Naqvi
Assistant Professor, Institute of Mathematics & Computer Science**,** University of Sindh, Jamshoro, Pakistan
Email: hira.naqvi@usindh.edu.pk

Dr. Zubair Ahmed Kalhoro
Associate Professor, Institute of Mathematics & Computer Science**,** University of Sindh, Jamshoro, Pakistan
Email: zubair.kalhoro@usindh.edu.pk

Muzaffar Hussain Laghari
Subject Specialist GBHSS Bhansinghabad, Mirpurkhas, Sindh Education and Literacy Department, Government of Sindh
Email: muzaffar.laghari@usindh.edu.pk

Abdul Rafiu Alias Furkan
Subject Specialist GBHSS Mithani, Naushahro Feroze, Sindh Education and Literacy Department, Government of Sindh
Email: fmemon022@gmail.com

ABSTRACT

Tuberculosis (TB) is one of the significant health issues of Pakistan. This paper reveals the long-term pattern of the TB cases reported in Pakistan over the period of 2002 – 2018 by the regression analysis. The time-based linear regression model to estimate annual TB data was done using Ordinary Least Squares (OLS). The study also projects the cases of TB until the year 2025 and determines the confidence bands to measure the uncertainty in the forecasts. The results show that there is a statistically significant increasing tendency in the occurrence of TB, and that it is high time to prevent the disease better, detect it at an early stage, increase its treatment compliance, and make informed changes in the policies. To visualize trends and uncertainty, a figure of actual and future cases with colored bands of uncertainty is also provided.

Key Words: Tuberculosis, Trend Analysis, Regression Analysis, Forecasting, Pakistan, Policy Recommendations

Introduction

One of the most common infectious diseases in the world is Tuberculosis (TB) which is a leading cause of morbidity and mortality, especially in developing countries (Qureshi et a1, 2025). World Health Organization (WHO) approximates that the number of people who are infected with the disease per year all over the world is millions and the disease has a tendency of targeting the areas that have weak health facilities and the ones that have social and economic imbalance (WHO, 2023). Pakistan is a country with a long history of being one of the high TB-burden nations with an important contribution to the global TB incidence rate. The national tuberculosis control programs and improved diagnostic abilities have not succeeded in reducing the number of reported TB cases in Pakistan indicating the current transmission and still existing challenges to the public health (NTP, 2025).

The importance of tracking and examining trends in TB incidence is not only to assess the effectiveness of the public health interventions but it is also used in making policy decisions. Epidemiological trends across time are best analyzed using statistical and mathematical modeling methods to gain a useful framework to study them. Linear regression is one of these, an easy but effective way to measure the trends of time, approximate the average changes every year and determine the statistical significance (Assuncao, 1993). Regression analysis enables the authorities in the field of public health to determine the periods of accelerated growth or decline and whether the interventions adopted have produced quantifiable effect on the trend in disease.

In addition to identifying trends, future cases of TB must be predicted so that there are proactive planning and allocation of resources. Predictive models will allow the policymakers to predict the possible growth of the number of cases and to distribute healthcare resources, diagnostic devices, and treatment courses more effectively. The use of confidence intervals or prediction bands in forecasts further measures the uncertainty level of the forecasts that makes the predictions an even stronger foundation in evidence-based decision-making (Wheelwright et al., 1998).

The socio-economic and demographic factors in Pakistan have a role to play in the increased cases of TB, whereas biological factors have a smaller role. Overpopulation, poverty, malnutrition, and the lack of good healthcare promote the existence and propagation of TB. The issue of urbanization and migration also makes the control of diseases more challenging, because any overcrowding of living conditions and poor sanitation are likely to lead to the spread. These contextual factors are crucial in the interpretation of TB trends and in the development of interventions that will focus on medical and social determinants of health.

Regression modeling and forecasting are data-based methods that are essential in contemporary policies in the field of public health. Quantification of trends and projecting future burden of diseases can enable policymakers make sound decisions about resource allocation, priorities of the programs to be implemented and measures to be taken. These methods are also useful to keep track of progress towards the national and international TB control goals, such as those set by the WHO End TB Strategy.

Objectives

This paper uses the linear regression analysis to explore the trend of TB-causing cases in Pakistan between the years 2002 and 2018 and then further makes predictions till the year 2025.

The analysis gives an empirical analysis of the past trends, estimates the number of additional cases of TB each year and quantifies the uncertainty of the forecasts using 95 percent confidence bands. The findings would be useful in terms of public health planning, policy development, and priorities of intervention measures to manage the current tuberculosis burden in Pakistan.

Research Questions

What is the trend of tuberculosis cases in Pakistan from 2002 to 2018?
Is the observed trend statistically significant based on regression analysis?
What are the forecasted TB cases in Pakistan up to 2025 and their associated uncertainty?

Literature Review

Quantitative modeling and statistical analysis have been playing a pivotal role in comprehending the dynamics of diseases and influencing the decisions made by populations on national health. The basis of the relationship analysis between epidemiological variables is the classical regression and econometric methods. (Assunicao ,1993) as well as Gujarati (1980) note that the linear regression models explain variation in observed data and determine the significance of the parameters, and this forms the basis of empirical analysis of the studies conducted in health and economics. The approaches are widely used in determining determinants of disease burden and testing the impacts of interventions. The time series analysis has become an effective method of predicting the patterns of infectious diseases. A detailed introduction to time series approaches is given by (Dickey & Chatfield, 1991), with great emphasis being given to stationarity, autocorrelation, and model diagnostics. Predictive techniques that are discussed by (Wheelwright et al., 1998) also emphasize the practical relevance of the model selection and predictive accuracy. Based on them, (Chen et al., 2013) use time series on pulmonary tuberculosis (TB) incidence data and show that time series can be effective in capturing patterns of incidence and give short term predictions that can be used in health planning. In addition to statistical forecasting, mathematical modeling has found extensive application in describing the dynamics of transmission of infectious diseases. The theoretical basis of compartmental models is laid by (Anderson and May, 1992) seminal work Infectious Diseases of Humans: Dynamics and Control which highlights the importance of reproduction numbers, stability studies and control measures. The concepts are still considered pivotal in the modern TB modeling research, with an insight into the mechanisms of transmission as the key to the assessment of disease control policy. International disease burden analyses offer much-needed background information to important modeling. Through the Global Burden of Disease Study (Murray & Lopez, 1997), they put a measure to mortality and disability that can be attributed to the major risk factors, thus showing that tuberculosis remains a global health problem. To supplement this, the reports of the World Health Organization (2023) offer recent estimates of TB incidence, prevalence, and mortality rates and describe strategic targets within the framework of the End TB Strategy. The reports are reliable sources of data and references in assessing model-based predictions. Recent publications pay more attention to the optimization and the comparison of mathematical models in order to enhance the predicting capabilities. (Qureshi et al., 2013) provides a comparative analysis of tuberculosis models, which reveals that the structure of models and parameterization have a considerable impact on the dynamics of the disease and the policy implications. These comparative methods are in line with current tendencies of focusing on robustness, sensitivity analysis, and evidence-based model selection. Correct diagnosis is an important element of TB control and modeling accuracy. The review by (Pai et al., 2017) shows the progress in TB diagnostics and the significance of diagnostic delays and sensitivity to the estimation of disease transmission. These are some diagnostic considerations which are important in interpretation of surveillance data that is being applied in statistical and mathematical models. Empirical data are essential at the national level in order to provide a context-specific analysis. The article (Qadeer et al., 2016) presents the initial population-based national TB prevalence survey in Pakistan, which gives credible estimates on disease burden among adults. The dataset has come to form a strong foundation in TB modeling research in Pakistan that has been relevant in calibration, validation, and policy analysis. On the whole, the literature shows that the combination of the econometric tools, time series forecasting, and mathematical modeling backed by high-quality surveillance and diagnostic data provides a powerful tool of analyzing the dynamics of tuberculosis. This approach of such integration is the key to proper forecasting, sound policy analysis and moving towards global objectives of TB elimination.

Data Description

Annual TB cases in Pakistan (2002–2018) were collected from World Health Organization website (World Health Organization, 2023) and converted into million. The dependent variable is the number of TB cases (I_t) in year t, and the independent variable is the time index (t), with $t = 1$ corresponding to 2002.

Methodology

Regression Model

To examine the temporal trend in tuberculosis infections, a simple linear regression model is specified as follows:

$$I_t = \beta_0 + \beta_1 t + \varepsilon_t \quad (1)$$

where I_t represents the number of TB-infected cases in year t, β_0 is the intercept (estimated number of cases at the initial time period), β_1 is the slope (average annual change in TB cases), and ε_t is the stochastic error term, capturing unobserved factors affecting TB incidence.

Ordinary Least Squares (OLS) Estimation

The parameters β_0 and β_1 are estimated using the Ordinary Least Squares (OLS) method, which minimizes the sum of squared differences between observed and predicted TB cases. The OLS estimators are given by:

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n}(t-\bar{t})(I_t-\bar{I})}{\sum_{t=1}^{n}(t-\bar{t})^2}, \quad \hat{\beta}_0 = \bar{I} - \hat{\beta}_1 \bar{t} \quad (2)$$

where n is the number of observations, $\bar{I}$ and $\bar{t}$ denote the mean of TB cases and time index, respectively.

Residuals and Error Analysis

The residuals are defined as the difference between observed and predicted values:

$$\varepsilon_t = I_t - \hat{I}_t \quad (3)$$

The residuals are used to calculate the standard error of the regression model:

$$s = \sqrt{\frac{\sum_{t=1}^{n}\varepsilon_t^2}{n-2}} \quad (4)$$

The standard error measures the typical deviation of observed TB cases from the predicted values and is used to construct confidence intervals for regression coefficients and

ISSN Online: 3007-3154
ISSN Print: 3007-3146

Vol. 4 No. 1 (January) (2026)
forecasts.

Correlation Coefficient Formula

The strength and direction of the linear relationship between time $t$and tuberculosis cases $I_t$are measured using the Pearson correlation coefficient r, defined as

$$r = \frac{\sum_{t=1}^{n}(t-\bar{t})(I_t-\bar{I})}{\sqrt{\sum_{t=1}^{n}(t-\bar{t})^2}\sqrt{\sum_{t=1}^{n}(I_t-\bar{I})^2}}$$

where

- I_t denotes the observed number of TB cases at time t,
- $\bar{I}$ is the mean of the observed TB cases,
- $\bar{t}$ is the mean of the time variable, and
- n is the total number of observations.

Coefficient of Determination (R^2)

The goodness of fit of the regression model is measured by the coefficient of determination, R^2, which quantifies the proportion of total variation in TB cases explained by the model:

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(I_t-\hat{I}_t)^2}{\sum_{t=1}^{n}(I_t-\bar{I})^2} \quad (5)$$

An R^2 value close to 1 indicates that most of the variability in TB cases is explained by the time variable.

Adjusted Coefficient of Determination

To account for the number of explanatory variables included in the regression model, the adjusted coefficient of determination is defined as

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-k-1}\right)\frac{\sum_{t=1}^{n}(I_t-\hat{I}_t)^2}{\sum_{t=1}^{n}(I_t-\bar{I})^2}$$

where

- $I_t$is the observed number of TB cases at time t,
- $\hat{I}_t$is the model-predicted number of TB cases,
- $\bar{I}$is the mean of observed TB cases,
- $n$is the total number of observations, and
- $k$is the number of explanatory variables in the regression model.

F-statistics for Overall Model Significance

To test whether the regression model explains a significant portion of the variation in TB cases, the F-statistic is computed:

$$F = \frac{\text{ExplainedVariation}/1}{\text{UnexplainedVariation}/(n-2)} = \frac{SSR/1}{SSE/(n-2)} \quad (6)$$

where SSR is the regression sum of squares and SSE is the sum of squared residuals. A high F-value (with a low p-value) indicates that the model is statistically significant.

t-test for Regression Coefficients

The statistical significance of each coefficient is tested using a t-test:

$$t_i = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \quad (7)$$

where $SE(\hat{\beta}_i)$ is the standard error of the estimated coefficient. The null hypothesis

$H_0: \beta_i = 0$ is rejected if the t-statistic exceeds the critical value from the t-distribution, indicating that the coefficient is significantly different from zero.

Forecasting TB Cases

Future TB cases are predicted using the estimated regression equation:

$$\hat{I}_{t+h} = \hat{\beta}_0 + \hat{\beta}_1(t+h) \tag{8}$$

where h represents the forecasting horizon. The 95% prediction interval for forecasted values is calculated as:

$$\hat{I}_{t+h} \pm t_{0.025,n-2} \cdot s\sqrt{1 + \frac{1}{n} + \frac{(t+h-\bar{t})^2}{\sum_{t=1}^{n}(t-\bar{t})^2}} \tag{9}$$

This takes care of model uncertainty as well as variability in the data giving the confidence ranges of the predicted TB cases.

Empirical Results

Estimated Regression Model

The fitted linear regression model using the TB data from 2002 to 2018 is obatined as:

$$\hat{I}_t = 0.0682 + 0.0190\, t \tag{10}$$

where $\hat{I}_t$ is the predicted number of TB-infected cases diuring the year t.

Regression Estimates and Statistical Measures

Table 2 summarizes the estimated regression coefficients, (including their standard errors, t-values, and p-values). Other model statistics including standard error of the residual, coefficient of determination (R^2) and F-statistic are also provided.

Table 2: OLS Regression Results

Parameter	**Estimate**	**Std. Error**	$t-value$	$p-value$
Intercept (β_0)	0.068227	0.014554	4.6879	< 0.00029
Time (β_1)	0.018994	0.001420	13.373	$< 9.7025 \times 10^{-10}$

Table 3: Statistical Measures

Statistic	Value
Residual Standard Error (s)	0.0287
Correlation	0.9605
Coefficient of Determination (R^2)	0.9226
Adjusted R^2	0.9175
F-statistic	$178.84 (p < 9.7025e^{-10})$

Confidence Intervals for Regression Coefficients

To compute the 95 percent confidence interval of the regression coefficients one calculates as:

$$\beta_i \pm t_{0.025,n-2} \cdot SE(\hat{\beta}_i) \tag{11}$$

- Intercept (β_0): $[0.037206,\ 0.099248]$
- Slope (β_1): $[0.015967,\ 0.022021]$

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

Vol. 4 No. 1 (January) (2026)

The confidence interval of β_1 neither includes the value of zero, so we can conclude that the upward trend in the number of TB-infected cases is statistically significant, $2002 - 2018$.

Interpretation of Results

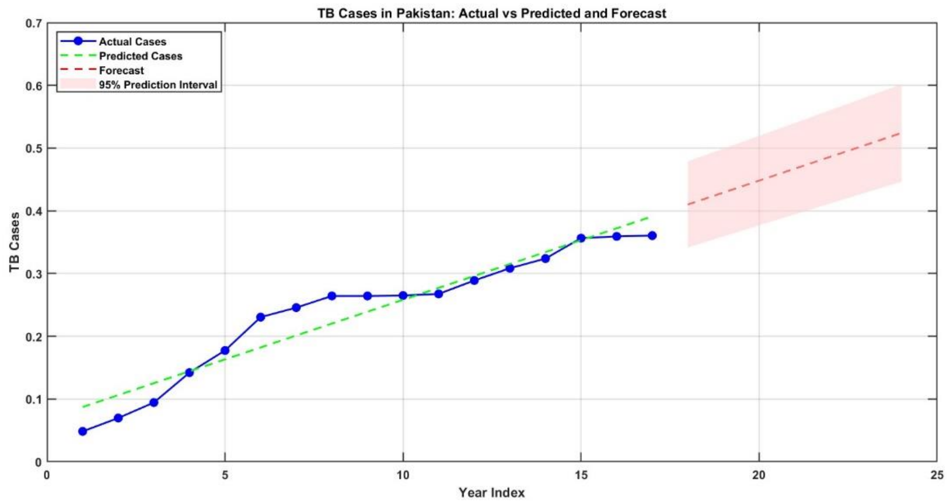
The slope coefficient $\hat{\beta}_1 = 0.0190$ indicates that, on average, the number of TB-infected cases increased by approximately 0.0190 cases per year during $2002 - 2018$. The high R^2 value (0.9226) confirms that the regression model explains most of the variability in TB cases, and the significant F-statistic ($p < 0.001$) validates the overall model significance. These results provide strong statistical evidence of a persistent upward trend in tuberculosis incidence in Pakistan.

Forecasting TB Cases (2019–2025)

Table 3: Forecasted TB Cases with Confidence Bands (2019–2025)

Year	Forecast	Lower Bound	Upper Bound
2019	0.41012	0.34155	0.47869
2020	0.42912	0.35929	0.49894
2021	0.44811	0.37692	0.51930
2022	0.46710	0.39445	0.53976
2023	0.48610	0.41188	0.56031
2024	0.50509	0.42923	0.58096
2025	0.52409	0.44649	0.60168

Figure 1: Trend of Actual TB Cases (2002–2018) and Forecasted Cases (2019–2025)



Discussion

The outcomes demonstrate that the number of cases with tuberculosis infection in Pakistan is increasing statistically significantly during the years $2002 - 2018$. This trend is attributable to various factors among them, high growth in the population, better case detection and better reporting systems. Also, the continued spread of TB especially in overpopulated and under-served regions has led to the increasing number of cases. The regression analysis offers a quantitative support that average cases of TB have been rising on average at a rate of almost 0.0190/per year. Projections of future cases to the year 2025 indicate that this trend is bound to increase unless there are specific measures to address the issue. The uncertainty and the possibility of the increased number of cases is additionally stressed by the 95 percentage intervals around the predictions. Such results demonstrate the growing population health issue of TB in Pakistan. Therefore, active measures such as diagnosis, better adherence to treatment and multifaceted programs to prevent tuberculosis are urgently needed to prevent the increase of this disease.

Policy Recommendations

This research has presented some recommendations on the ways to successfully curb and mitigate tuberculosis burden in Pakistan based on the results of the present study. To begin with, the early TB detection programs are to be enhanced, and the access to the modern diagnostic facilities is to be increased, especially in rural and high-risk regions. It is important to ensure that effective treatment protocols are followed and this can be done through patient-centered care, periodic follow-ups and surveillance systems to avoid default of treatment and subsequent development of drug-resistant strains. The level of awareness creation and community involvement should be enhanced by educating the population about TB prevention, early symptoms, the necessity of adherence to treatment, and reducing social stigma of the disease.

The government should increase its investments in TB surveillance, case reporting, and healthcare infrastructure to help to monitor it properly and intervene in time. The use of forecasting and statistical modeling tools in the national TB programs will help in the evidence based planning and efficient use of resources. Cooperation among the public health bodies, non governmental organizations as well as international organizations should be encouraged to exchange knowledge, best practices and new ways of handling TB. Moreover, novel TB vaccines, diagnostic tools and treatment solutions should be researched and developed so as to fight the drug resistant strains. The medical personnel must be provided with continuous education to enhance diagnosis of TB, management and counseling of patients. The interventions should be targeted at high-burden cities and vulnerable groups, such as migrants, slum residents, and malnourished people. Lastly, a periodic review of TB programs and policies through data-driven methods will ensure the identification of gaps, strategies optimization, and sustainable long-term decrease in the incident of tuberculosis.

Conclusion

This article shows that TB cases in Pakistan $(2002 - 2018)$ statistically show an upward trend and will increase to 2025 with computed confidence bands. These results highlight the importance of specific prevention, early diagnosis, proper treatment, and informed policy interventions to reduce the burden of TB. Regression analysis and forecasting have a quantitative basis of evidence-based decision-making and resource allocation. The active contributions of healthcare authorities, policymakers, and community stakeholders are the conditions to reduce the spread of TB and obtain the long-term results of improving the population health in the country.

References

Assuncao, R., Sampson, P. D., Montgomery, D. C., & Peck, E. A. (1993). Introduction to linear regression analysis. *Journal of the American Statistical Association*, *88*(421), 383. https://doi.org/10.2307/2290746

Chen, Y. P., Wu, A. P., Wang, C. L., Zhou, H. Y., & Feng, S. X. (2013). Time series Analysis of pulmonary tuberculosis incidence: Forecasting by applying the Time Series

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 4 No. 1 (January) (2026)
model. *Advanced Materials Research*, *709*, 819–822.
https://doi.org/10.4028/www.scientific.net/amr.709.819

Dickey, D. A., & Chatfield, C. (1991). The Analysis of Time Series: An Introduction. *Technometrics*, *33*(3), 363.
https://doi.org/10.2307/1268794

Wheelwright, S., Makridakis, S., & Hyndman, R. J. (1998). *Forecasting: methods and applications*. John Wiley & Sons.
https://doi.org/10.1016/0305-0548(93)90040-p

Gujarati, D. (1980). *Basic Econometrics, 4th Edition*.
https://openlibrary.org/books/OL17075482M/Basic_econometrics

Anderson, R. M., & May, R. M. (1991). *Infectious diseases of humans: dynamics and control*. Oxford university press.
https://doi.org/10.7326/0003-4819-117-2-174_4

Murray, C. J., & Lopez, A. D. (1997). Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *The Lancet*, *349*(9063), 1436–1442.
https://doi.org/10.1016/s0140-6736(96)07495-8

National TB Control Programme (NTP) – Common Management Unit. (n.d.). Accessed on 02/12/2025.
https://www.cmu.gov.pk/ntp-national-tb-control-programme/

Qureshi, Z. A., Qureshi, S., Shaikh, A. A., & Shahani, M. Y. (2025). Optimizing tuberculosis dynamics through a comparative evaluation of mathematical models. *Communications in Mathematical Biology and Neuroscience.*
https://doi.org/10.28919/cmbn/9248

Pai, M., Nicol, M. P., & Boehme, C. C. (2017). Tuberculosis Diagnostics: State of the art and future directions. In *ASM Press eBooks* (pp. 361–378).
https://doi.org/10.1128/9781555819569.ch16

Qadeer, E., Fatima, R., Yaqoob, A., Tahseen, S., Haq, M. U., Ghafoor, A., Asif, M., Straetemans, M., & Tiemersma, E. W. (2016). Population Based National Tuberculosis Prevalence Survey among Adults (>15 Years) in Pakistan, 2010–2011. *PLoS ONE*, *11*(2), e0148293.
https://doi.org/10.1371/journal.pone.0148293

World Health Organization, "Global tuberculosis programme data," 2023, Accessed: 02/12/2025.
[Online]. Available: https://www.who.int/teams/global-tuberculosis-programme/data