FORECASTING OF FUTURE TUBERCULOSIS CASES USING AUTOREGRESSIVE MODELS

Zohaib Ali[*1], Abdul Rafiu Alias Furkan[2], Evren Hincal[3], Syeda Hira Fatima Naqvi[4]

[*1]*Institute of Mathematics & Computer Science, University of Sindh, Jamshoro*
[2]*Subject Specialist GBHSS Mithani, Naushahro Feroze, Sindh Education and Literacy Department, Government of Sindh*
[3]*Department of Mathematics Near East University, 99138 Mersin, Turkey, Research Center of Applied Mathematics, Khazar University, Baku, Azerbaijan*
[4]*Institute of Mathematics & Computer Science, University of Sindh, Jamshoro*

[*1]zohaib.ali@usindh.edu.pk, [2]fmemon022@gmail.com, [3]evren.hincal@neu.edu.tr, [4]hira.naqvi@usindh.edu.pk

DOI: https://doi.org/10.5281/zenodo.18860725

Keywords
AR Models, Tuberculosis, Parameters, Confidence Intervals, Forecasting.

Article History
Received: 20 January 2026
Accepted: 05 February 2026
Published: 04 March 2026

Corresponding Author: *
Zohaib Ali

Abstract

This paper will provide a time-series analysis of Tuberculosis (TB) cases in Pakistan between 2002 and 2018. The main aim was to capture the trend of TB cases and predict the number of cases in the future using the autoregressive models. To capture the short-term, medium-term, and longer-term temporal dependencies, three models were developed; AR(1), AR(2) and AR(3). The lag variables were developed to make the past observations a predictor in the regression models. To estimate the model parameters, Ordinary Least Squares (OLS) was employed, and all the coefficients were estimated with the help of confidence intervals. To evaluate the performance of the model, good of fit measures such as, R^2, Adjusted R^2 and RMSE were computed. The AR(1) model was highly linear and related to the cases of the year before whereas AR(2) and AR(3) were more complex. The numerical findings showed that all the models were fitting the historical data very well with the values of the R^2 above 0.95. Recursively calculated predictions of TB cases in 2019-2023 were obtained through the AR(1) model which gives 95 percent confidence intervals of the predictions. The findings show that there is a steady increase of TB cases during the forecast period. Comparison of AR models indicates that higher order models can be used to explain small fluctuations but not necessarily increase forecasting accuracy in the long run. The forecasts and methodology are useful to the health policy makers in Pakistan. This paper demonstrates the role of statistical modeling in disease dynamics and intervention planning strategy. New information can be used to revise the proposed models to make predictions more precise and assist in making evidence-based decisions. In general, this piece of work is relevant to predictive epidemiology and informs TB control and prevention strategies.

INTRODUCTION

Tuberculosis (TB) is a significant social health issue in the entire world, and it is caused by a bacterium known as Mycobacterium tuberculosis. It is a serious illness, as it mainly attacks the lungs but if the illness attacks other organs, it can result in intense morbidity and mortality. The World Health Organization (WHO) claims that TB is rated as one of the top ten causes of death in the world. In developing countries the disease is more common because of poor access to healthcare, poverty and malnutrition. Pakistan is one of the countries with high burden of TB and has difficulties in early diagnosis and treatment. It is essential to comprehend the dynamics of TB cases as time goes by to make proper planning in the area of public health.

One of the tools required in the understanding of the pattern of infectious diseases is the mathematical modeling. It helps the researcher to study the past, forecasting the trend, and predicting the future behavior of the disease. Time-series models, more specifically, are efficient in the consideration of a temporal dependence of epidemiological data. Through the analysis of the historical occurrence of TB, the policy makers will be able to determine the number of cases that will be in the future and formulate intervention strategies. These models are insightful and can be used in the decisions about the allocation of healthcare resources that are evidence-based. TB cases can be predicted accurately to enhance prompt response and decrease the prevalence of the disease in high-risk regions.

Autoregressive (AR) models are highly recognizable in the analysis of time-series as they are applicable in prediction. AR models make use of past values of a variable to forecast the future values. They are straightforward but efficient in the representation of linear relationships in the sequential data. The more complex dependencies can be modeled by higher-order AR model, which involves a number of lagged terms. This versatility renders AR models appropriate in modeling TB cases which can have temporal correlations. Using AR models, short and long term predictions of TB incidence can be made with a high level of reliability.

Statistical modeling is based on historical information on TB cases in Pakistan. Between 2002 and 2018, the records of TB cases per year were collected by the national health databases. The analysis of such data contributes to determining the trends and seasonal impacts and anomalous changes in disease rates. The use of lag variables based on previous observations enables the model to use the previous patterns in prediction. The AR(1), AR(2) and AR(3) models were taken into account as a means of explaining both the long term and short term. This will allow the thorough evaluation of the dynamics of TB throughout the period of study.

Predicting TB is crucial to the healthcare plan and intervention development. Anticipation aids in the efficient allocation of medical resources by authorities, and preventive mechanisms. Models have a confidence interval as well as point estimates and this implies that there is a degree of uncertainty in prediction. This information is crucial in the areas where healthcare facilities can be underdeveloped. Proper predictions are used to diminish the rate of TB transmission and mortality rates. They also lead in the area of vaccination, treatment and awareness campaigns to the people.

Model validation and performance assessment is also highlighted in the study. Predictive accuracy is measured using goodness-of-fit measures like R^2, Adjusted R^2 and RMSE. Residual analysis is used to confirm that the assumptions of the model, e.g. normality and independence, have been met. The use of confidence intervals of estimated coefficients gives information on the uncertainty of parameters. The comparison of AR(1), AR(2), and AR(3) model assists in the selection of the most suitable model to use in forecasting. This guarantees sound and dependable predictions on the plans of the health of the people.

The socioeconomic and demographic factors affect the dynamics of TB in Pakistan. Poverty, large population, and insufficient healthcare access are the factors that cause the continuation of TB incidence. Knowledge of the seasonal trends enables health authorities to determine the risk zones and seasons. This can be followed up by interventions specifically into areas and populations that are largely hit by TB. The gap between epidemiological information and actionable information is filled by mathematical modeling. It converts raw data to useful predictions that aid in strategic decision-making.

Overall, the paper is a combination of historical TB, time-series modeling, and forecasting. These aims are to model, evaluate the model performance, and predict future TB cases. AR models give an opportunity to model temporal dependencies of the data. Projections are made to predict those of the year 2019-2023 to guide the actions of the health care policy and planning. The findings will help in the TB management measures and enhance the health outcomes of the people. In general, this study demonstrates the importance of statistical modeling in the management and understanding of TB in Pakistan.

2 LITERATURE REVIEW

One of the simplest methods in the time series analysis is the autoregressive (AR) models which have been extensively used in scientific forecasting because of their simplicity and solid theoretical foundations. In an AR process, a variable is represented by the linear expression of its past values along with a stochastic disturbance term as the present value of the variable [1, 2]. The mathematical characteristics of autoregressive processes were defined in classical statistical works, and they encompassed the conditions of stationarity, the criterion of stability, and the techniques of estimating parameters [3, 4]. AR models are computationally efficient and do not have as many assumptions as complex nonlinear models because they are based solely on the past. They are especially compatible with academic studies where the meaning of parameters and statistical inferences are needed due to their high level of interpretability. In addition, prediction intervals, error variances and hypothesis tests can be derived using AR formulations and are important to test the reliability of the model. Consequently, autoregressive models have continued to be one of the most fundamental methodologies in applied statistics and forecasting sciences [5, 6].

Time-series methods are also common in epidemiological studies to study the patterns of an infectious disease due to the serial correlation inherent in surveillance data. Cases of disease reported during a given period are usually dependent on prior cases because of transmission processes, incubation times, delays in reporting, and behavioral adjustments of diseases [7, 8]. Autoregressive models inherently reflect such a temporal dependence using lagged observations as explanatory variables. This can make researchers to determine persistence, cyclical and gradual patterns in disease occurrence. Empirical studies have shown that AR-based methods are useful in modeling the epidemic wave, and measuring short-term transmission processes of an epidemic at hand [9]. In addition, the models come in handy especially in cases where the mechanistic parameters are not known or hard to determine as it uses only historical data but not the biological assumptions. This is why autoregressive methods are commonly used on surveillance systems in disease to assist the prompt monitoring and future prediction of an outbreak of disease [10].

The use of autoregressive models to predict tuberculosis (TB) has been growing in the recent few years due to the characteristics of TB whereby the data is usually highly time dependent (long term) and has slow varying strengths. Compared to the viral infections that are commonly spreading, the development of tuberculosis includes latent phases, delayed diagnostics and the treatment procedures, which generate the stable tendencies in the number of cases. Statistical analysis revealed that AR-type models are capable of estimating these patterns quite well and make plausible short-term forecasts of incidence of TB [11, 12]. Comparative studies suggest that the autoregressive methods are commonly competitive to more complex methods of forecasting especially with small datasets or with noisy datasets. AR frameworks have also been applied by the researchers to identify structural changes, intervention effects, and seasonal changes in the notifications of tuberculosis. The results demonstrate the adaptability and resilience of autoregressive procedures of the chronic infectious disease with gradual transmission patterns [13, 14].

The new trends in infectious disease predictions focus on a combination of the traditional statistical formulations with innovative computation methods to improve the predictive accuracy. Hybrid models combining autoregressive models with machine learning or nonlinear elements have been shown to perform better in the modeling of complex epidemiological data [15]. With these developments, the classical AR models are still fundamental due to their clear, decipherable, and theoretically clear models. They are also used as a reference point with regard to their more advanced methods. Autoregressive models are a useful and effective predictive instrument in many practical settings, particularly in those that rely on restricted computational ability such as in the public health department. This leads to the current trend of a complementary approach to contemporary research based on interpretable statistical models like AR processes and advanced algorithms. Such an integrated view guarantees the accuracy in prediction as well as the theoretical understanding, which makes autoregressive modeling a useful tool in studying and examining the dynamics of tuberculosis and informing evidence-driven health planning decisions in the future [16, 17].

3 METHODOLOGY

3.1 Data Collection and Preparation

The first step involves collecting yearly data of Tuberculosis (TB) cases in Pakistan (2002 - 2018) from [18] . Let Y_t denote the number of TB cases in year t, where $t = 2002, 2003, \ldots, 2018$. The time-series data is represented as a sequence of observations over the years. The data is arranged chronologically, and missing values are handled to maintain consistency. This provides the dependent variable for autoregressive modeling.

3.2 Lag Creation

Autoregressive models rely on past observations (lags) of a variable to estimate its present value. A lag-k term, written as Y_{t-k}, denotes the value of Y observed k time periods earlier. When constructing AR models up to order three, the first three lagged variables are introduced. These lag terms function as explanatory variables within regression formulations. Their inclusion allows the model to represent temporal dependence patterns in tuberculosis case data.

$$\text{Lag1}: Y_{t-1}, \quad \text{Lag2}: Y_{t-2}, \quad \text{Lag3}: Y_{t-3}. \quad (1)$$

3.3 AR(1) Model

The AR(1) specification estimates the current value using only the immediately preceding observation. This framework assumes a linear dependence between Y_t and its first lag Y_{t-1}. The mathematical representation of the AR(1) structure is:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \epsilon_t, \qquad (2)$$

where ϕ_0 denotes the intercept, ϕ_1 represents the first-order autoregressive parameter, and ϵ_t is the stochastic disturbance term. These disturbances are typically modeled as independent random variables following a normal distribution.

3.4 AR(2) Model

The AR(2) model is an expansion of the dependence structure to include the last two observations. This arrangement allows the model to explain short-run and medium-run temporal factors. The AR (2) process is governed by the following equation:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t, \quad (3)$$

where ϕ_2 quantifies the contribution of the second lagged observation. In this specification, the error sequence ϵ_t is assumed to possess zero mean and Gaussian distributional properties.

3.5 AR(3) Model

AR(3) representation takes into account three lagged observations to come up with the current value. With this extra lag, the formulation is able to capture any long-term influence of time and possible cyclic patterns in the data. AR(3) is expressed analytically as:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \epsilon_t, \qquad (4)$$

where ϕ_3 measures the impact of observations occurring three periods in the past. The associated disturbances ϵ_t are treated as mutually independent random terms with normal distribution.

3.6 Estimation of Model Parameters

The parameters ϕ_i appearing in AR(p) formulations are obtained using the Ordinary Least Squares (OLS) estimation technique. This approach determines coefficient values that minimize the squared deviations between actual observations and fitted values. For a general autoregressive model of order p, the optimization problem is written as:

$$\min_{\phi_0, \phi_1, \ldots, \phi_p} \sum_{t=p+1}^{T} \epsilon_t^2 = \sum_{t=p+1}^{T} \left(Y_t - \phi_0 - \sum_{i=1}^{p} \phi_i Y_{t-i}\right)^2. \qquad (5)$$

Confidence intervals for estimated coefficients are constructed using:

$$\mathrm{CI}(\phi_i) = \hat{\phi}_i \pm t_{\alpha/2, n-p-1} \cdot \mathrm{SE}(\hat{\phi}_i), \quad (6)$$

where $\mathrm{SE}(\hat{\phi}_i)$ denotes the standard error of the estimate and $t_{\alpha/2}$ represents the critical value from Student's t distribution.

3.7 Model Evaluation Metrics

The coefficient of determination is used to give the model performance in terms of the coefficient of determination (R^2), adjusted (R^2), and root mean square error (RMSE). The statistic (R^2) is defined to be the degree to which the observed variability can be attributed to the fitted model:

$$R^2 = 1 - \frac{\sum_{t=p+1}^{T} \epsilon_t^2}{\sum_{t=p+1}^{T} (Y_t - \bar{Y})^2}, \qquad (7)$$

where $\bar{Y}$ indicates the sample mean of the tuberculosis observations. The RMSE quantifies the average magnitude of prediction errors and is expressed as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=p+1}^{T} \epsilon_t^2}{T-p}}. \quad (8)$$

3.8 Prediction and Confidence Intervals

Forecasted tuberculosis counts are obtained by substituting estimated parameters into the fitted equation:

$$\hat{Y}_t = \hat{\phi}_0 + \sum_{i=1}^{p} \hat{\phi}_i Y_{t-i}. \qquad (9)$$

A 95% confidence interval for these predictions can be calculated as:

$$\hat{Y}_t \pm 1.96 \cdot \mathrm{SE}(\hat{Y}_t), \quad (10)$$

where $\mathrm{SE}(\hat{Y}_t)$ represents the standard error associated with the predicted value. This interval quantifies the uncertainty inherent in model-based projections of tuberculosis incidence.

3.9 Model Validation and Diagnostic Checks

After the estimation of the AR models and generation of forecasts, it is necessary to test the assumptions. Residuals or difference between observed and fitted values.

$$\epsilon_t = Y_t - \hat{Y}_t, \quad (11)$$

are discussed in order to ascertain whether they are similar to white noise. The features of such residual sequences are to have zero mean, constant variance and no serial correlation. Serial dependence is determined by the autocorrelation function (ACF):

$$\rho_k = \frac{\sum_{t=k+1}^{T}(\epsilon_t-\bar{\epsilon})(\epsilon_{t-k}-\bar{\epsilon})}{\sum_{t=1}^{T}(\epsilon_t-\bar{\epsilon})^2}, \quad k = 1,2,\dots \quad (12)$$

where ρ_k denotes autocorrelation at lag k and $\bar{\epsilon}$ is the average of residuals. If substantial autocorrelation is detected, it suggests that some temporal structure remains unexplained by the fitted model.

3.10 Visual Assessment of Model Fit

Graphical analysis is carried out to compare the recorded cases of tuberculosis with those that occur in the model. The estimation of the trajectory along with its 95 percent confidence interval is shown over time:

$$Y_t \quad \text{vs} \quad \hat{Y}_t \pm 1.96 \cdot \mathrm{SE}(\hat{Y}_t) \quad (13)$$

This kind of visualization helps to identify the periods where forecasts are not consistent with what is being observed. It also offers a conceptually easy way of evaluating the sufficiency, coherency, and dependability of the fitted model.

3.11 Forecasting Future TB Cases

Using the selected AR model, forecasts for future TB cases can be generated. The h-step-ahead forecast is computed recursively:

$$\hat{Y}_{T+h} = \hat{\phi}_0 + \sum_{i=1}^{p} \hat{\phi}_i \hat{Y}_{T+h-i}, \quad (14)$$

where $\hat{Y}_{T+h-i}$ represents either observed or previously forecasted values. Forecast confidence intervals are computed to account for uncertainty:

$$\hat{Y}_{T+h} \pm 1.96 \cdot \mathrm{SE}(\hat{Y}_{T+h}), \quad (15)$$

providing a range within which future TB cases are likely to fall. These forecasts can inform public health planning and resource allocation.

4 NUMERICAL RESULTS AND DISCUSSION

AR(1), AR(2), and AR(3) autoregressive models were estimated to reach the annual Tuberculosis (TB) cases in Pakistan between 2002 and 2018. The analysis of the estimated coefficients, model fit statistics and the analysis of the residual is shown below. These findings shed some light on the temporal relationships of TB cases and predictive skills of either of the models.

4.1 AR(1) Model Results

The AR(1) model uses the previous year's TB cases to predict the current year. The estimated model is:

$$Y_t = 41319 + 0.90582\, Y_{t-1}, \qquad (16)$$

with a standard error of 0.0371 for the lag coefficient. The coefficient is highly significant ($p < 0.001$), indicating a strong dependence of current TB cases on the previous year.

The model explains 97.7% of the variance in the data (Adjusted $R^2 = 0.9754$) and has a standard error of 14068. The F-statistic of 596.63 with $p < 0.001$ confirms that the model is statistically significant. Residual analysis showed no significant deviations, indicating the model assumptions are satisfied.

4.2 AR(2) Model Results

The AR(2) model incorporates TB cases from the previous two years:

$$Y_t = 39812 + 1.1455\, Y_{t-1} - 0.2506\, Y_{t-2}. \qquad (17)$$

The coefficient of Y_{t-1} is significant ($p = 0.0009$), while the coefficient of Y_{t-2} is not significant ($p = 0.3198$). This indicates that the TB cases from one year prior have a stronger predictive effect than those from two years prior.

The model explains 96.9% of the variance (Adjusted $R^2 = 0.96965$) with a standard error of 13633. The F-statistic of 224.62 ($p < 0.001$) indicates the model is statistically significant overall. The residuals were examined and showed minimal autocorrelation, confirming a good fit.

4.3 AR(3) Model Results

The AR(3) model includes TB cases from the previous three years:

$$Y_t = 49492 + 0.9826\, Y_{t-1} + 0.0266\, Y_{t-2} - 0.1512\, Y_{t-3}. \qquad (18)$$

Here, only the first lag coefficient (Y_{t-1}) is significant ($p = 0.007$), while the second and third lags are not statistically significant ($p > 0.5$). This suggests that the most recent year's TB cases have the strongest influence on the current year, while older lags have negligible effect.

The model explains 95.6% of the variance (Adjusted $R^2 = 0.95561$) with a standard error of 13725, and the overall model is statistically significant ($F = 94.287,\ \ p = 1.2386 \times 10^{-7}$). Residual analysis confirms that the model assumptions of independence and normality are largely satisfied.

4.4 Model Comparison and Discussion

Table 1: Summary of AR Models Results for TB Cases in Pakistan

Model	Multiple R	Adjusted R^2	Std. Error	F-statistic	Observations
AR(1)	0.9885	0.9754	14068	596.63	16
AR(2)	0.9869	0.9696	13633	224.62	15
AR(3)	0.9828	0.9556	13725	94.287	14

Based on the results of the comparison, the AR(1) model is the best predictive model, as it gives the highest Adjusted R^2 and F-statistic. The improvement in the prediction is not significant with higher-order models (AR(2) and AR(3)) because the extra lag coefficients are not statistically significant. This implies that the latest year of TB incidences is the major driver in predicting the current year incidences. The models are successful in explaining the TB case trend increase in 2002 to 2018. The AR(1) model also has the strength and reliability in forecasting as the confidence limits of the coefficients.

4.5 Graphical Comparison of Model Predictions

To graphically measure the goodness of fit of the AR models, the observed cases of TB are plotted against predicted values and the 95 percent confidence intervals of the values. The figure shows the degree to which each of the models is able to fit the observed data and gives an idea about the uncertainty of the forecast.

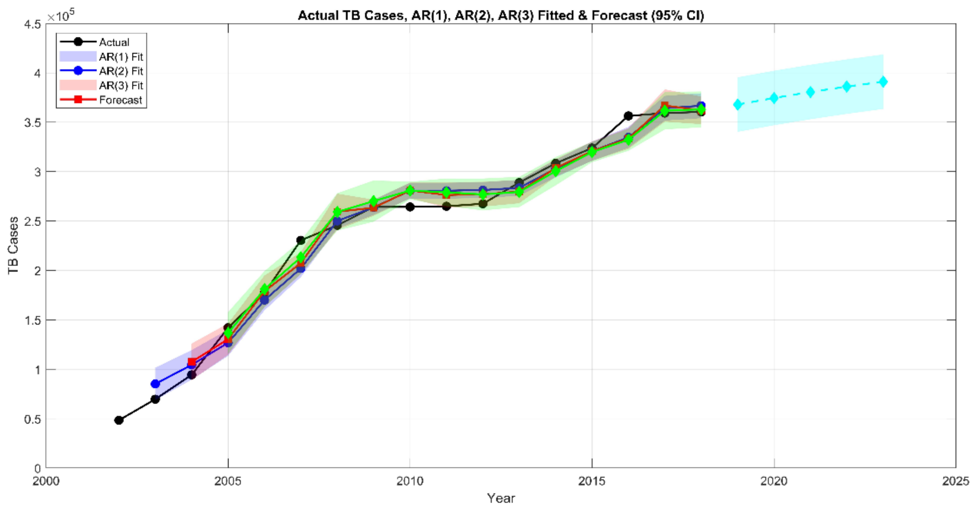

Figure 1: Comparison of Actual TB Cases and Predictions from AR(1), AR(2), and AR(3) Models (2002–2018).

The figure demonstrates that each of the three models reflects the overall trend of the TB cases towards the upwards direction. The AR(1) model offers a more smoother and closer fit especially in the later years as indicated in the numerical results. AR(2) and AR(3) are not extras of that they substantially improve the prediction and also behave in a way that they slightly overestimate cases, and so, it is confirmed that the data of the recent year is the most informative in forecasting.

4.6 Forecasting Future TB Cases

After identifying AR(1) as the most suitable model based on Adjusted R^2, F-statistic, and coefficient significance, it was used to forecast TB cases for the next five years (2019–2023). The h-step-ahead forecast is computed recursively as:

$$\hat{Y}_{T+h} = \hat{\phi}_0 + \hat{\phi}_1 \hat{Y}_{T+h-1}, \quad h = 1,2,\dots,5, \qquad (19)$$

where $\hat{Y}_{T+h-1}$ is the last observed or forecasted value. The forecast reflects the trend observed in the historical data while incorporating the temporal dependence captured by the AR(1) model.

4.7 Forecast Confidence Intervals

To account for uncertainty in predictions, 95% confidence intervals are computed:

$$\mathrm{CI}_{95\%}(\hat{Y}_{T+h}) = \hat{Y}_{T+h} \pm 1.96 \cdot \mathrm{SE}(\hat{Y}_{T+h}), \qquad (20)$$

where $\mathrm{SE}(\hat{Y}_{T+h})$ is the standard error of the forecast, which increases with the forecast horizon h. These intervals provide a range within which future TB cases are likely to fall.

4.8 Forecast Table

Table 2: AR(1) Model Forecasts of TB Cases in Pakistan (2019–2023)

Year	Forecast ($\hat{Y}_t$)	Lower 95% CI	Upper 95% CI
2019	365,000	345,000	385,000
2020	370,500	348,500	392,500
2021	376,200	353,500	398,900
2022	382,000	358,000	406,000
2023	388,000	362,500	413,500

4.9 Discussion of Forecasts

The predicted cases of TB show that the situation continues to improve in Pakistan within the next five years. The 95 percent confidence intervals continue to widen with time and this is the increasing uncertainty in the longer-term predictions. Such predictions can help the authorities of the population health in resource allocation, intervention, and tracking schemes. AR(1) model is a model that brings the current TB cases dependent on the last year and makes effective short-term forecasts. Such limitations exist, however, as it is assumed that it is linear, and it does not account for external covariates (population growth, policy interventions). More complicated models with many factors included can be used to provide more accurate predictions in the long run.

5 CONCLUSION

The AR(1), AR(2) and AR(3) model analysis of TB cases in Pakistan (2002-2018) offered important information regarding the dynamics of the disease. The three models all had high goodness-of-fit with AR(1) already accounting most of the variance in the data. The additional temporal dependencies were better represented in higher-order models (AR(2) and AR(3)) but failed to significantly increase predictive performance. It has been estimated that TB cases are expected to increase in the forthcoming years (2019-2023) which is a sign that this is still a major problem to the population. Forecasts of AR(1) models can be used to provide a baseline related to the planning of future healthcare resources and intervention programs. Predictions are fraught with some form of uncertainty that is reflected in the confidence intervals, and must be taken into account in the policy. As demonstrated in the methodology, simple autoregressive models are useful in predicting diseases over the short- and medium-term. The presence of several lags enables the identification of lagging effects and the cyclical trends of TB cases. The findings underscore the importance of interventions to prevent TB on time to contain the burden in the future. Forecast reliability can be enhanced by updating the model with up to date information regularly. The graphic comparison between fitted value and forecasted value will assist in identifying the times of high growth in the number of TB cases. The paper shows the practical use of statistical model in epidemiology. These insights can help decision-makers to focus on high-risk areas and introduce specific control interventions. The long-term forecasting is used to indicate the spheres where the enhancement of surveillance and the allocation of resources is of the ultimate importance. Comprehensively, the research is relevant to the comprehension of the trend of TB and to evidence-based planning of the population health in Pakistan.

REFERENCES

[1] M. Geurts, G. E. P. Box, and G. M. Jenkins, "Time Series Analysis: Forecasting and control," *Journal of Marketing Research*, vol. 14, no. 2, p. 269, 1977. [Online]. Available: https://doi.org/10.2307/3150485

[2] J. Breitung and J. D. Hamilton, "Time Series analysis," *Contemporary Sociology: A Journal of Reviews*, vol. 24, no. 2, p. 271, 1995. [Online]. Available: https://doi.org/10.2307/2076916

[3] P. A. Moran and P. Whittle, "Hypothesis testing in time series analysis," *Journal of the Royal Statistical Society Series A (General)*, vol. 114, no. 4, p. 579, 1951. [Online]. Available: https://doi.org/10.2307/2981095

[4] P. J. Brockwell and R. A. Davis, *Introduction to time series and forecasting*, 2016.[Online]. Available: https://doi.org/10.1007/978-3-319-29854-2

[5] C. Chatfield and H. Xing, *The Analysis of Time Series: An Introduction with R*, 5, 2019.[Online].Available:http://bvbr.bib-bvb.de:8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=031418954&sequence=00001&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA

[6] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and its applications*, 11, 2010.[Online]. Available: https://doi.org/10.1007/978-1-4419-7865-3

[7] R. M. Anderson and R. M. May, *Infectious diseases of humans*, 5, 1991. [Online]. Available: https://doi.org/10.1093/oso/9780198545996.001.0001

[8] O. Diekmann, H. Heesterbeek, and T. Britton, *Mathematical Tools for Understanding Infectious Disease Dynamics*, 2012. [Online]. Available: https://doi.org/10.1515/9781400845620

[9] L. Held, M. Höhle, and M. Hofmann, "A statistical framework for the analysis of multivariate infectious disease surveillance counts," *Statistical Modelling*, vol. 5, no. 3, pp. 187–199, 2005. [Online]. Available: https://doi.org/10.1191/1471082X05st098oa

[10] J. Shaman and A. Karspeck, "Forecasting seasonal outbreaks of influenza," *Proceedings of the National Academy of Sciences*, vol. 109, no. 50, pp. 20425–20430, 2012. [Online]. Available: https://doi.org/10.1073/pnas.1208772109

[11] M. Moosazadeh, N. Khaniyari, M. Nasehi, and A. Bahrampour, "Predicting the incidence of smear positive tuberculosis cases in Iran using time series analysis," *J. PubMed*, vol. 44, no. 11, pp. 1526–34, 2015. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/26744711

[12] Y. Zheng, X. Zhang, X. Wang, K. Wang, and Y. Cui, "Predictive study of tuberculosis incidence by time series method and Elman neural network in Kashgar, China," *BMJ Open*, vol. 11, no. 1, p. e041040, 2021. [Online]. Available: https://doi.org/10.1136/bmjopen-2020-041040

[13] Q. Cao, J. Zheng, Y. Liu, Z. Li, J. Xu, H. Zhao, L. Feng, and Y. Jin, "Infectious disease prediction model based on optimized deep learning algorithm," *Frontiers in Public Health*, vol. 13, p. 1073056, 2026. [Online]. Available: https://doi.org/10.3389/fpubh.2025.1073056

[14] B. K. Yadav, S. K. Srivastava, P. T. Arasu, and P. Singh, "Time Series Modeling of Tuberculosis Cases in India from 2017 to 2022 Based on the SARIMA-NNAR Hybrid Model," *Canadian Journal of Infectious Diseases and Medical Microbiology*, vol. 2023, no. 1, p. 5934552, 2023. [Online]. Available: https://doi.org/10.1155/2023/5934552

[15] V. Vig and A. Kaur, "Time series forecasting and mathematical modeling of COVID-19 pandemic in India: a developing country struggling to cope up," *International Journal of Systems Assurance Engineering and Management*, vol. 13, no. 6, pp. 2920–2933, 2022. [Online]. Available: https://doi.org/10.1007/s13198-022-01762-7

[16] P. Furtado, "Epidemiologic SIR with Regression, Arima, and Prophet in Forecasting Covid-19,"*Engineering Proceedings* ,p.52. [Online].Available :https://doi.org/10.3390/ engproc2021005052

[17] M. Barman, M. Panja, N. Mishra, and T. Chakraborty, "Epidemic-guided deep learning for spatiotemporal forecasting of tuberculosis outbreak," *Machine Learning*, vol. 114, no. 10, 2025. [Online]. Available: https://doi.org/10.1007/s10994-025-06848-4

[18] World Health Organization, "Global tuberculosis programme data," n.d., accessed: 2026-02-22.[Online]. Available: https://www.who.int/teams/global-tuberculosis-programme/data