Abstract

Background: Pulmonary tuberculosis (PTB) remains major public health problem over the world. Cities witnessing religious event throughout of the year like Kerbala/Iraq require great efforts to minimize the incidence of deadly communicable diseases like TB. The aim of this study is to model the monthly incidence rates of PTB cases in Kerbala/Iraq. Methods: This is a retrospective study in which records of confirmed PTB patients whom they referred to the chest and respiratory illnesses center of Holy Kerbala governorate were obtained. Monthly registered new smear-positive PTB cases from January 2010 to December 2016 were analyzed. Seasonal autoregressive integrated moving average (SARIMA), SARIMA-exponential smoothing method (ETS), SARIMA-neural network autoregressive, and SARIMA-adaptive neuro-fuzzy inference system (SARIMA-ANFIS) were used for forecasting monthly incidence rate of TB in Kerbala, Iraq. Mean absolute percentage error, root mean square error, and mean absolute square error were used to compare the models, and Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used to select best model. Results: The trend of PTB incidence showed a seasonal characteristic, with peaks in spring and winter. Predicted estimates using all models proposed to forecast the number of PTB cases from 2016 to 2018 showed that the PTB cases indicated marginal decrease trends and best forecasted in SARIMA-ANFIS model (the lower AIC and BIC values, 712.69 and 731.05, respectively). Conclusion: Seasonal characteristic of PTB incidence was observed with peaks during spring and winter. Forecasting of PTB incidence between the period 2016 and 2018 showed marginal decrease trends, and the best forecasting model was SARIMA-ANFIS model.

Keywords: Forecasting incidence, pulmonary tuberculosis, seasonality

Introduction

Pulmonary tuberculosis (PTB) caused by the bacteria Mycobacterium tuberculosis complex (MTB) remains major global health threat in spite of the great efforts for controlling this communicable disease. Approximately 10.4 million infected people with TB were diagnosed in 2016, and 2 million patients died. This disease occurs through two pathways, transmission or reactivation of latent infection. Several previous retrospective studies described the presence of seasonal pattern in PTB incidence. Some of them reported that the highest incidence was seen during late spring or early summer. High PTB case notification in spring and summer may suggest that the infection is peak in winter taking into account the period of disease appearance and diagnosis. The reason behind seasonality is still unknown. However, there are two hypotheses explaining it. First, Vitamin D deficiency hypothesis which considered as immune regulator factor that improves immunity against MTB infection in vitro. Subsequently, decreased exposure to the sun during winter months may lead to vitamin deficiency which might increase the likelihood for endogenous reactivation of PTB latent infection. The other hypothesis declared that increased PTB incidence might result from increase transmission rather than activation during wintertime indoor crowding conditions. Iraq with other eastern Mediterranean region (EMR) countries accounts for 25% of the global PTB incidence in 2014. In
addition, it is considered among the nine high TB burden countries in the EMR with estimated 20000 PTB cases and 4000 deaths annually contributing to 3% of the total cases.[2] Moreover, according to the Iraqi Ministry of Health Report 2012, the estimated PTB cases in Iraq was 45/100,000.[12]

Kerbala is one of the governorates of Iraq and is the second holiest cities in Iraq after Al-Najaf in Shia Islam. This governorate witnesses mass-gathering events every year and thus should be in focus, especially for communicable diseases including PTB. Although seasonal variation of PTB infection has been reported in several previous studies in different countries, no previous study was conducted to assess seasonal patterns of PTB cases in Holy Kerbala/Iraq. Understanding seasonal variation in this holy region may help in understanding the transmission and pathogenesis of PTB, recognizing risk determinant, and improve PTB control and management. The aim of the current study is to model the seasonal trends and forecast incidence rates of PTB infections in Holy Kerbala/Iraq.

**METHODS**

This is a retrospective study in which records of confirmed PTB patients whom they referred to the chest and respiratory illnesses center of Holy Kerbala governorate between the period January 2010 and December 2016 were obtained after taking approval from relevant committees belong to the Ministry of Health/Iraq.

These records were documented using the criteria of the World Health Organization/International Union of Tuberculosis and Lung Diseases.[13] The date of diagnosis including day, month, quarter, and year was investigated for assessing the seasonal patterns of the disease. Up to 602 newly diagnosed PTB cases were reported during this period.

**Forecasting models**

Three forecasting comparison methods were used for forecasting monthly TB incidence from 2010 to 2016; seasonal autoregressive integrated moving average (SARIMA), SARIMA-exponential smoothing method (SARIMA-ETS), SARIMA-neural network autoregressive (SARIMA-NNAR), and SARIMA-adaptive neuro-fuzzy inference system (SARIMA-ANFIS). The “forecast” packages in R software with version 5.0.1 were used to fit the SARIMA model. The “caret” and “NNAR” packages in R software with version 5.0.1 were used to fit the SARIMA-NNAR model. The SARIMA-ANFIS model was done using “ANFISR” in R package, version 5.1 to fit the model. Autocorrelation function (ACF) and partial ACF (PACF) were also investigated.

**Neural network model**

The feedforward neural network with backpropagation training algorithm in transferring function in the single hidden layer of the tan-sigmoid function expressed as follows:

\[
n_n = f(n_i) = \frac{1 - e^{-n_i}}{1 + e^{-n_i}}, \quad i = 1, 2, 3, \ldots, P
\]

where \( n_i \) is the neurons of the input vector and \( P \) is the number of neurons. We used Levenberg–Marquardt algorithm to training a function for minimization of the network performance function. This algorithm incorporates Newton’s algorithm and the gradient descent to increase the tentative step performance function and will shifts toward Newton’s method if the reduction of the performance function is successful. However, each iteration of the algorithm used to reduce the performance function. To prevent over-fitting problem, we used Bayesian regularization and early stopping methods to improve the network generalization. The methods used to provide a measure of how many network parameters are being effectively used by the network. The artificial neural network (NNAR) widely used as forecasting methods that give nonlinear association between dependent and independent variables in training algorithm for the multilayer feedforward network. The NNAR is used to store number of input-output mapping patterns without expressing its mathematical model to show the relationship between inputs and outputs. The network studies the input-output relationship gradually by adjusting the weights and thresholds to minimize the error between the observations and predictions. The process can involve three layers at least as follows: input layer that receives and distributes input patterns, middle or hidden layer that captures the linearity or nonlinearity of the input–output relationship patterns, and output layer that produces the output fitted results.

**Adaptive neuro-fuzzy inference system model**

The ANFIS model consists of two components: a fuzzy inference system and a back propagation algorithm. The learning method works similarly to that of neural networks. The fuzzy inference incorporated into the ANFIS is the first-order Sugeno-type inference, the typical rule of which, if there are only two inputs \( x \) and \( y \), and the output \( z \) has the form.

\[
Z = \frac{\sum_{i=1}^{N} w_i z_i}{\sum_{i=1}^{N} w_i}
\]

Where \( N \) is the number of the rules.

For SARIMA-ANFIS combine forecast proposed was derived using the fitted values derived from the SARIMA model as inputs in the neural network, and the actual data of PTB cases were the outputs to improve the prediction performance of the conventional SARIMA model.

**Model comparison and evaluation**

Three models were adopted to fit and predict the monthly incidence of PTB. Mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute square error (MASE) were used to compare the models, and Akaike information criterion (AIC) was used to selected best model.

\[
MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{\hat{x}_i - x_i}{x_i} \right| \times 100\%
\]
Mohammed, et al.: Seasonality of PTB

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (\hat{x}_t - x_t)^2}
\]

\[
\text{MASE} = \frac{1}{N} \sum_{t=1}^{N} \left( \frac{1}{N-1} \sum_{i=t}^{N} |\hat{x}_i - x_i| \right)
\]

When comparing models fitted by likelihood criteria to the data, the smaller the AIC or Bayesian information criterion, the better the fit. The theory of AIC requires that the log-likelihood has been maximized to give the best fit.

**RESULTS**

Among 3254 patients referred to the chest and respiratory illnesses center, 602 confirmed to have new and active PTB. The data were recorded between the period January 2010 and December 2016. The male/female ratio was 1.13, and the mean age ± standard deviation was 44.02 ± 17.23, as provided in our previous study. The overall PTB cases during the studied period were decreased. A time series plot was drawn to identify the trend and pattern of the TB dataset over the period of years. Figure 1 indicates the behavioral pattern of monthly and annual TB incidence of the 7-year period.

SARIMA is an ARIMA model consider as a seasonal trend in a time series data. SARIMA model was transformed into a stationary time series by trend and seasonal difference and later fit the data. After checking the data for seasonality adjustment, the data were seasonally decomposed using the additive method to identify the seasonality components, patterns, and effects of the seasonal adjusted values, Figure 1 shows the seasonally adjusted values. The seasonal adjusted values were used to observe the PTB occurrence cases with a specific seasonal difference trend. The trend of PTB incidence showed a seasonal characteristic, with peaks in spring and winter.

We established the differences between ACF and PACF. They are both showing if there is a significant correlation between a point and lagged points. The difference is that PACF takes into consideration the correlation between each of the intermediate lagged points. Looking at ACF could be misleading with what points are significant, that is, there is a strong correlation, as shown in Figure 2.

**Forecast from the models**

Table 1 shows the summary of the model comparison and evaluation for the goodness of fit. The measures of errors in MAPE, RMSE, and MASE were observed to be lower in SARIMA-ANFIS model compared to other models. The best model was selected with the lowest AIC (712.69).

We predicted that the estimates using all the four models proposed to forecast the number of TB cases yearly from 2016 to 2018 are summarized in both Table 2 and Figures 3-5. It is observed that the TB cases indicated marginal decrease trends and best forecasted in SARIMA-ANFIS model.

**Prediction interval for single model**

Presumably, we used mean prediction intervals rather than confidence intervals. The fitted values are in-sample one-step forecasts. We assumed that the fitted values are normally distributed errors with 95% prediction intervals given by \( YT \pm 1.96 \sigma^2 \) where \( \sigma^2 \) is the estimated variance of the residuals. A prediction interval is used in a forecast combination. We assumed that the forecast errors are uncorrelated and normally distributed, and then, a simple 95% prediction interval for the next observation in a time series were calculated and expressed in Figures 6-8.
A prediction interval is a similar but not identical concept to a confidence interval. A prediction interval is an estimate of a value that is going to be observed at some point in the future. A prediction interval of the individual randomness associated with the particular point or points being predicted.

**Table 1: Models comparison and evaluation**

| Model comparison       | RMSE | MAPE | MASE | AIC         | BIC         |
|------------------------|------|------|------|-------------|-------------|
| SARIMA model           | 59.34| 98.19| 0.76 | 922.26      | 929.52      |
| SARIMA-ETS model       | 58.53| 93.74| 0.75 | 1065.86     | 1078.02     |
| SARIMA-NNAR model      | 53.10| 90.41| 0.44 | 810.33      | 817.62      |
| SARIMA-ANFIS model     | 40.89| 81.87| 0.35 | 712.69      | 731.05      |

RMSE: Root mean square error, MAPE: Mean absolute percentage error, MASE: Mean absolute square error, AIC: Akaike information criterion, BIC: Bayesian information criterion, ETS: Exponential smoothing method, NNAR: Neural network autoregressive, ANFIS: Adaptive neuro-fuzzy inference system.

Figures 6-8 show prediction intervals calculated to include the true results 95% of the time to estimate the range of values conditional on the model being correct in the first place. It showed that the prediction interval for the forecast from SARIMA-ANFIS was the best within the estimated interval of 80%–85%.

The point estimate for prediction interval observed that the estimate for prediction interval for forecast from SARIMA-ANFIS has the best forecast prediction interval for the forecast accuracy for the PTB cases in the forecast years [Figures 9-13].

**Discussion**

Holy Kerbala represents the second holiest city in Iraq after Al-Najaf. In Kerbala several religious events occur throughout the year during which millions of pilgrims are
visiting Kerbala making this city requires great attention to minimize the transmission of communicable diseases like Tuberculosis. Several previous studies reported the presence and absence of seasonal pattern in the incidence of PTB cases. However, the presence or absence of seasonal variation has not described previously in Holy Kerbala/Iraq. Thus, the current study aimed to determine the seasonal pattern of PTB infection if present and predict the future values (using different models) which may enhance controlling and management of the infection. Time series plot identified behavioral pattern of monthly and annual PTB cases incidence of the studied period. The data were seasonally decomposed using the additive method to identify the seasonality components, patterns, and effects of the seasonal adjusted values. The seasonal adjusted values were used to observe the PTB occurrence cases with specific seasonal difference trends. The trend of PTB incidence showed a seasonal characteristic, with peaks in spring and winter.

The presence of spring peak was reported in several previous studies. Inversely, the presence of summer peak was reported in other previous studies. Ríos et al. showed the peak of TB cases in summer and autumn quarters. Seasonal peak was reported in spring and summer in another study in the north of Iran. However, the real cause behind seasonal patterns of...
PTB remains unknown. Each season involves various conditions such as weather, length of daylight, and others. Several previous studies showed that cold weather and shorter sunlight in winter cause Vitamin D deficiency and this subsequently leads to weakness of the immune response and increase endogenous reactivation of PTB incidence.[9,19] Moreover, indoor air pollution, poorly ventilated housing conditions, and higher time spent indoor might increase the transmissibility of MTB which could be seen in the winter season.[20]

The weather in Iraq during winter months is wet and cold, and thus, many people may get viral and bacterial respiratory infections, and this may contribute to lowering the immune response and subsequently increase vulnerability to PTB infection.[21] In contrast, summer months in Iraq are hot and dry and thus much less people get respiratory infection. Furthermore, during winter months’ decrease sun exposure due to short daylight and stay indoor could enhance transmission of MTB. Inversely, summer months in Iraq are very hot and represent the holiday months for most of the people thus people try to minimize sun exposure which may lead to Vitamin D deficiency (which represent as immune modulator) and increase susceptibility to PTB infection either through endogenous reactivation of bacteria or through susceptibility to new infection. Vitamin D deficiency had been reported recently in the Kerbala population[22] and the most important source for Vitamin D dermal synthesis-dependent upon sunlight exposure. The minimum exposure to sunlight during summer and winter could lead to vitamin deficiency and subsequent impairment of immunity and minimize the macrophage ability for intracellular killing of bacteria.[23]

Time series modeler was applied using SARIMA, SARIMA-ETS, SARIMA-NNAR, and SARIMA-ANFIS for forecasting monthly incidence rate of TB in Holy Kerbala, Iraq. Forecasting the future incidence of PTB from 2016 to 2018, we predicted the estimates using all the four models proposed, and the results observed that the PTB cases indicated marginal decrease trends and best forecasted in SARIMA-ANFIS model.

Moreover, the current study established the differences between ACF and PACF. They are both showing if there is a significant correlation between a point and lagged points. Looking at ACF could be misleading with what points are significant, that is, there is a strong correlation.

**Conclusion**

Seasonal characteristic of PTB incidence was observed with peaks during spring and winter. Forecasting of PTB incidence between the period 2016 and 2018 showed marginal decrease trends, and the best forecasting model was SARIMA-ANFIS model.

**Financial support and sponsorship**

Nil.

**Conflicts of interest**

There are no conflicts of interest.

**References**

1. Azeez A, Obarmi D, Odeyemi A, Ndege J, Muntabayi R. Seasonality and trend forecasting of tuberculosis prevalence data in Eastern Cape, South Africa, using a hybrid model. Int J Environ Res Public Health 2016;13. pii: E757.
2. World Health Organization. Global Tuberculosis Report 2016. World Health Organization; 2016.
3. Willis MD, Winston CA, Heilig CM, Cain KP, Walter ND, MacKenzie WR, et al. Seasonality of tuberculosis in the United States, 1993-2008. Clin Infect Dis 2012;54:1553-60.
4. Khalid A, Batooli SA, Chaudhry MN. Seasonality and trend analysis of tuberculosis in Lahore, Pakistan from 2006 to 2013. J Epidemiol Glob Health 2015;5:397-403.
5. Naranbat N, Nymadawa P, Schopfer K, Rieder HL. Seasonality of tuberculosis in an Eastern-Asian country with an extreme continental climate. Eur Respir J 2009;34:921-5.
6. Wubuli A, Li Y, Xue F, Yao X, Upur H, Wushouer Q, et al. Seasonality of active tuberculosis notification from 2005 to 2014 in Xinjiang, China. PLoS One 2017;12:e0180226.
7. Coussens A, Timms PM, Boucher BJ, Venton TR, Ashcroft AT, Skolimowska KH, et al. 1alpha, 25-dihydroxyvitamin D3 inhibits matrix metalloproteinases induced by Mycobacterium tuberculosis infection. Immunology 2009;127:539-48.
8. Davies PD. A possible link between Vitamin D deficiency and impaired host defence to Mycobacterium tuberculosis. Tubercle 1985;66:301-6.
9. Wilkinson RJ, Llewelyn M, Toossi Z, Patel P, Passvol G, Lalmvani A, et al. Influence of Vitamin D deficiency and Vitamin D receptor polymorphisms on tuberculosis among Gujarati Asians in West London: A case-control study. Lancet 2000;355:618-21.
10. Thorpe LE, Frieden TR, Laserson KF, Wells C, Khatri GR. Seasonality of tuberculosis in India: Is it real and what does it tell us? Lancet 2004;364:1613-4.
11. Schaaf HS, Nel ED, Beyers N, Gie RP, Scott F, Donald PR, et al. A decade of experience with *Mycobacterium tuberculosis* culture from children: A seasonal influence on incidence of childhood tuberculosis. Tuber Lung Dis 1996;77:43-6.

12. Iraq Ministry of Health. New Diagnosis of Tuberculosis in Iraq. Baghdad, Iraq: In: Tech. Rep., Ministry of Health of Iraq; 2012.

13. Munir MK, Anwar N, Iqbal R, Shabbir I, Nosheen S. Diagnosis of tuberculosis: Molecular versus conventional method. Pak J Med Res 2011;50:50.

14. Sharonova A, Loza K, Surmeneva M, Surmenev R, Prymak O, Epple M. Synthesis of Positively and Negatively Charged Silver Nanoparticles and Their Deposition on the Surface of Titanium. In: IOP Conference Series: Materials Science and Engineering. IOP Publishing; 2016.

15. Douglas AS, Strachan DP, Maxwell JD. Seasonality of tuberculosis: The reverse of other respiratory diseases in the UK. Thorax 1996;51:944-6.

16. Leung CC, Yew WW, Chan TY, Tam CM, Chan CY, Chan CK, et al. Seasonal pattern of tuberculosis in Hong Kong. Int J Epidemiol 2005;34:924-30.

17. Ríos M, García JM, Sánchez JA, Pérez D. A statistical analysis of the seasonality in pulmonary tuberculosis. Eur J Epidemiol 2000;16:483-8.

18. Moosazadeh M, Khanjani N, Bahrampour A. Seasonality and temporal variations of tuberculosis in the north of Iran. Tanaffos 2013;12:35-41.

19. Liao CM, Hsieh NH, Huang TL, Cheng YH, Lin YJ, Chio CP, et al. Assessing trends and predictors of tuberculosis in Taiwan. BMC Public Health 2012;12:29.

20. Douglas AS, Ali S, Bakhshi SS. Does vitamin D deficiency account for ethnic differences in tuberculosis seasonality in the UK? Ethn Health 1998;3:247-53.

21. Narula P, Sihota P, Azad S, Lio P. Analyzing seasonality of tuberculosis across Indian states and union territories. J Epidemiol Glob Health 2015;5:337-46.

22. Al-Hilali K. Prevalence of hypovitaminosis D in adult iraqi people including postmenopausal women. Sci Res J 2015;4:53-62.

23. Koh GC, Hawthorne G, Turner AM, Kunst H, Dedicoat M. Tuberculosis incidence correlates with sunshine: An ecological 28-year time series study. PLoS One 2013;8:e57752.