SHORT REPORT
doi: 10.2306/scienceasia1513-1874.2009.35.392

ScienceAsia 35 (2009): **392–395**

Tuberculosis in the Terengganu region: Forecast and data analysis

Adem Kilicman[a,*], **Ummu Atiqah Mohd Roslan**[b]

[a] Department of Mathematics, Faculty of Science, University Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

[b] Department of Mathematics, Faculty of Science and Technology, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia

[*]Corresponding author, e-mail: akilicman@science.upm.edu.my

Received 16 Dec 2008
Accepted 6 Nov 2009

ABSTRACT: In this study we analyse tuberculosis (TB) data from Jabatan Kesihatan Negeri Terengganu (2008) by applying linear trend, quadratic trend, simple moving average, simple exponential smoothing and Holt's trend corrected exponential smoothing. Accuracy of these time series approaches are measured by computing the variance between the extrapolation model and the actual data. The study shows that Holt's trend corrected exponential smoothing is the best forecasting model, followed by the quadratic trend model. The results also show that people aged between 35–44 years old, male, Malay, unemployed or have an income lower than RM 1000 per month are in a high risk group to be infected by TB. We also forecast TB cases for 2009 until 2013 and the result suggests that the numbers of TB cases are expected to increase.

KEYWORDS: time-series forecasting, data analysis

INTRODUCTION

Tuberculosis (TB) is a common and deadly infectious disease caused by mycobacteria, mainly *Mycobacterium tuberculosis*. Tuberculosis usually attacks the lungs (pulmonary TB), but it can also affect the central nervous system, the lymphatic system, the circulatory system, the genitourinary system, bones, joints, and even the skin. The World Health Organization declared TB as a global health emergency in 1993, and developed a Global Plan to stop TB with the aim of saving 14 million lives between 2006 and 2015. In Malaysia, TB was the number one cause of death in the early 1940s and 1950s. In 2000, there were a total of 15 057 cases of all forms of TB notified in Malaysia. The state with the highest disease burden was Sabah followed by Wilayah Persekutuan, Sarawak, and Pulau Pinang. The majority of TB cases occur in the 15–54 years age group (67.7% in the year 2000). Human immunodeficiency virus (HIV) infection is the single most important risk factor for the development of active TB. The number of TB cases with HIV infection has also increased. In 1998, TB was the 11th leading killer disease in Malaysia [1].

In Sabah, a state in Malaysia, the average incidence of TB among health care workers over the past 5 years is 280.4 per 100 000 of the population. The research reported that the main risk factors of TB among health care workers are age, gender, history of TB contact outside the workplace (other than family contact), duration of service and failure to use respiratory protection when performing high-risk procedures [2]. For the treatment strategy we refer to Ref. 3. In 2000, Malaysia achieved a detection rate of smear positive cases of 70% and the cure rate was 77.6% [1]. This means that the case detection for the year 2000 was the same as the global target which was also 70% but the cure rate was only 77.6%. Therefore, it appears that there is still room for improvement, particularly in terms of achieving a better cure rate.

In 2006 Terengganu had a population of 1 080 286, of which Malays make up 94.7% of the population, Chinese, 2.6%, while Indians 0.2% and other ethnic groups comprise the remaining 2.4%. In 2000, 48.7% of the population lived in urban areas. By 2005 this had changed to 51%.

The objectives for this study are to draw the pattern of the data for TB cases in the Terengganu region over time in order to determine which forecasting model would most accurately predict TB cases in the Terengganu region, and also to identify the high risk group to be infected with TB and to forecast the number of TB cases after 2009. This study is based on a study of multi-drug resistant (MDR-TB) in Finland [4]. For analysing data, we will consider TB patients registered in Kuala Terengganu district only.

All of these patients are referred to Kuala Terengganu Chest Clinic (KTCC), which is situated near Sultanah Nur Zahirah Hospital.

METHOD

Time series forecasting techniques

We used time series analysis to forecast the future distribution of TB cases. Time series analysis is a quantitative method used to determine patterns in data collected regularly over time. We project these patterns to estimate future values and cope with uncertainty about the future.

For the forecasting, we obtained the data for the Terengganu region from Jabatan Kesihatan Negeri Terengganu. The data were from 1987 to 2006 which means that there were 20 data points. We can describe a time series y_t by using a trend model. The equation for the linear trend model is

$$y_t = \beta_0 + \beta_1 t. \tag{1}$$

In this model, β_1 represents the average change from one period to the next and is also known as the slope of the linear trend. The quadratic trend model is given by

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2. \tag{2}$$

In the simple moving average (SMA) technique, one can assume that the data patterns as exhibited by the historical observations can be best represented by an arithmetic mean or average of past observations. This technique requires revising the average as new data becomes available. The revision is done by incorporating the new observation in the group and dropping the last observation in the preceding group of values. The results of the calculations are the series of values that move from one time period to the next, over the entire period covered by the data series[5]. The SMA has a fixed length of the underlying moving average[6]. The SMA is represented by

$$\mathrm{SMA}_n = \frac{1}{n} \sum_{t=k-n+1}^{k} P_t, \quad k = n, \ldots, 20,$$

where n is the number of periods included in the average, k is the relative position of the period currently being considered within the total number of periods, and P_t is the number of TB cases at time t. It is easy to calculate moving average by using an odd number of periods because the values of moving averages are correctly placed at the centre of the affected time periods. Normally, 3-year SMA or 7-year SMA is used for this kind of model[5]. In this study, we choose $n = 3$ because the TB data we have obtained is non-seasonal data and the number of data points is small.

The simple exponential smoothing (SE) method is also used for forecasting a time series when there is no trend or seasonal pattern, but the mean of the time series y_t is slowly changing over time[7]. A single exponential smoothing method smooths the data by computing exponentially weighted averages and provides short-term forecasts. The model requires only one parameter, the smoothing constant α (where $0 < \alpha < 1$), in order to generate the fitted value and hence forecast[5]. The SE method is often used to forecast the value of the next observation, given the current and prior values. Thus if we already know the present value y_n then we try to forecast the value y_{n+1}. The model equation for SE is given by

$$S_t = \alpha X_t + (1 - \alpha) S_{t-1}, \tag{3}$$

where S_t is the forecasted number of TB cases, X_t is the number of TB cases for the year before, and S_{t-1} is the previous forecast.

Holt's trend corrected exponential smoothing (HT), also known as double exponential smoothing, smooths data and provides short-term forecasts. This procedure can work well when a trend is present but it can also serve as a general smoothing method and the dynamic estimates are calculated for the level and trend. Thus HT employs a level component and a trend component at each period. It uses two weights, or smoothing parameters, to update the components at each period[7]. The HT model equations are

$$\begin{aligned} S_t &= \alpha X_t + (1 - \alpha)(S_{t-1} - T_{t-1}) \\ T_t &= \gamma (S_t - S_{t-1}) + (1 - \gamma) T_{t-1} \\ \hat{X}_t &= S_{t-1} + T_{t-1} \end{aligned} \tag{4}$$

where S_t is the level at time t, α is the value of the smoothing parameter for the level, T_t is the trend at time t, γ is the value of the smoothing parameter for the trend, X_t is the data value at time t, and $\hat{X}_t$ is the fitted value, or one-period-ahead forecast, at time t. The values of α and γ were given by the MINITAB software. HT uses the level and trend components to generate forecasts.

To determine which forecasting model provides the best prediction of TB cases, we use only 15 data points. The remaining five data points are used for comparing with the forecast by computing the variance between the actual and forecast data. These processes were done for all time-series techniques in order to find the model which has the smallest variance.

Data Analysis

The second part of the study is to analyse the data by using SPSS software where retrospective analysis was done on TB patients who were registered in KTCC from January 2006 to December 2006 and the data were obtained from Pejabat Kesihatan Daerah Kuala Terengganu[8]. There were 645 patients with TB in the state of Terengganu. These were latent infections rather than symptomatic cases. For data analysis, we only considered 187 patients as all of them were from Kuala Terengganu district only. This is because the data we obtained for Kuala Terengganu are more detailed than for other districts. The remaining 64 patients were diagnosed in other six districts which are in Marang, Dungun, Kemaman, Kuala Berang, Setiu, and Besut.

The factors studied were socio-demographic characteristics (age, gender, ethnicity and employment status), place of residence (sub-district), income, type of TB, TB category, HIV, sputum smear for acid-bacilli (AFB), and the presence of pulmonary tuberculosis with or without extra-pulmonary involvement. The following age groups were analysed: 0–4, 5–14, 15–24, 25–34, 35–44, 45–54, 55–64, and above 65 years old. The four kinds of ethnicity included in this research were Malay, Chinese, Indian, and Indonesian. The employment status for TB patients are divided into employed and unemployed. Monthly income levels were divided into the following four categories: inconsistent, lower than RM 500, RM 500–999, RM 1000–3000, and more than RM 3000. The term 'inconsistent income' refers to unemployed people or self-employed people operating a small business.

The HIV status is either positive or negative, as is the Sputum AFB smear. Finally, pulmonary tuberculosis is classified as being with or without extra-pulmonary involvement. TB data was divided into four categories: new, recurrent, after default treatment, and treatment stopped.

All 10 variables were used as stated above and frequency table analysis and crosstabs analysis were carried out. In frequency table analysis, the SPSS software gives us the frequency and percentage tables. From this output, we determined which groups had the highest and the lowest frequency. In crosstabs analysis, chi-squared tests are used to test the difference between categorical factors studied and the relationship between the variables. The output gives the significant values for the relationships and therefore we can obtain the important factors that can be related to TB.

Table 1 TB Data for Terengganu Region.

Year	TB	Year	TB	Year	TB	Year	TB
1987	393	1993	323	1999	456	2005	595
1988	413	1994	373	2000	531	2006	645
1989	406	1995	401	2001	485	2007	663
1990	422	1996	366	2002	482	2008	730
1991	438	1997	444	2003	565		
1992	362	1998	504	2004	549		

Source: Jabatan Kesihatan Negeri Terengganu

RESULTS AND DISCUSSION

Time series forecasting techniques

The data values are given in Table 1. These include values for 2007 and 2008 obtained recently. By using MINITAB, we obtained the values of the parameters of the five time-series analysis techniques automatically. For the linear trend model (1), $\beta_0 = 363.019$, $\beta_1 = 7.26429$, thus indicating a positive trend. For the quadratic trend model (2), $\beta_0 = 440.257$, $\beta_1 = -19.9962$, and $\beta_2 = 1.7037$. For the SE model (3), $\alpha = 0.699$. For the HT model (4), $\alpha = 0.678$ and $\gamma = 0.157$.

Table 2 shows the actual TB data for Terengganu region and the predicted TB by using all the five techniques. These results were obtained by using MINITAB software. We also predict the number of TB cases for 2002–2006. It can be seen that HT is the best forecasting model since it has the smallest variance (Table 2). This is followed by SE, Quadratic Trend, Linear Trend, and SMA.

Five-year TB forecasting

We now use the HT model to generate the TB forecast for year 2009 until 2013. In order to forecast for five years forward, we use all the 22 data points in Table 1 and therefore in (4) the smoothing parameter values change: $\alpha = 0.629$ and $\gamma = 0.181$.

Fig. 1 shows the pattern of TB cases together with the predicted and forecast TB by using the MINITAB software. From Fig. 1, we can see that the number of TB forecast is expected to increase for the next five years.

Data analysis

The results show that the persons who are aged between 35–44 years old, male, Malay, unemployed and have income lower than RM 1000 are in a high risk group to be infected by TB. Of the 187 patients selected for the analysis, most were Malay (96.8%), 129 cases (69%) were male, 107 of them (57.2%) were unemployed, and most of them (59.3%) had

Table 2 Linear Trend, Quadratic Trend, 3-year SMA, SE and HT forecasts.

Period (year)	Actual TB	Linear Forecast	Quadratic Forecast	SMA Forecast	SE Forecast	HT Forecast
2002	482	479.25	556.49	556.49	492.70	511.64
2003	565	486.51	592.71	592.71	485.22	523.40
2004	549	493.78	632.35	632.35	540.99	535.17
2005	595	501.04	675.39	675.39	546.59	546.93
2006	645	508.30	721.85	721.845	580.43	558.69
	Variance	7346.33	5126.36	7831.90	2611.39	2511.89

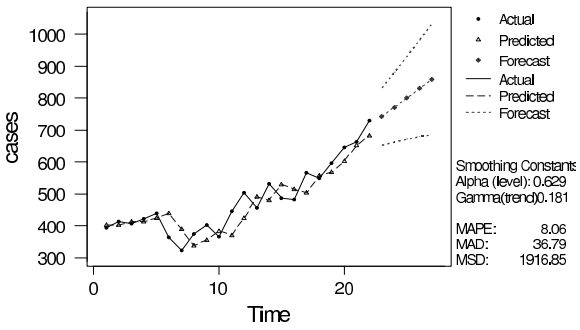

Fig. 1 Actual TB data (dots), the predicted value (squares), and the 5-year forecast (diamonds) by using HT model. Triangular points show the upper and lower bounds.

an income lower than RM 1000 per month, 19.8% between RM 500 and RM 999, 16.6% between RM 1000 and RM 3000, and the remaining 4.3% had an income of more than RM 3000. This shows that the number of TB cases decreases with increasing income. Loytonen and Maasilta[4] have also shown that income and age were significantly related to TB in Finland.

All cases came from Kuala Terengganu District. The three sub-districts with the most cases were Bandar (11.8%), Batu Rakit (13.4%) and Kuala Nerus (13.9%). There was only one case (0.5%) in each of the sub-districts Atas Tol, Bukit Tunggal, Kedai Buloh, and Pengadang Baru. Most of the cases (94.1%) were new tuberculosis cases, 105 cases (56.1%) had positive sputum smear for AFB, 39 cases (20.9%) had extra-pulmonary involvement, and 18 cases (9.6%) were HIV positive out of 187 patients.

In crosstabs analysis, we found four significant relationships. These were between TB with age and gender ($P = 0.013$), between TB with gender and employment status ($P = 0.000$), between TB with gender and HIV ($P = 0.014$), and between TB with employment status and income ($P = 0.000$).

Acknowledgements: The authors thank the referee for the very constructive report and suggestions that increase the quality of the paper very much. The authors would also like to thank the Hospital Tuanku Nur Zahirah Record Office (Kuala Terengganu) for providing the data to make this analysis. The authors also acknowledge that the research is partially supported by Academic Fellowship Scheme at University of Malaysia Terengganu.

REFERENCES

1. Kuppusamy I (2004) Tuberculosis in Malaysia: problems and prospect of treatment and control. *Tuberculosis* **84**, 4–7.
2. Jelip J, Mathew GG, Yusin T, Dony JF, Singh N, Ashaari M, Lajanin N, Ratnam CS, Ibrahim MY, Gopinatha D (2004) Risk factors of tuberculosis among health care workers in Sabah, Malaysia. *Tuberculosis* **84**, 19–23.
3. Raviglione MC (2003) The TB epidemic from 1992 to 2002. *Tuberculosis* **83**, 4–14.
4. Loytonen M, Maasilta P (1998) Multi-drug resistant tuberculosis in Finland – a forecast. *Soc Sci Med* **46**, 695–702.
5. Mohd Alias Lazim (2005) *Introductory Bussiness Forecasting – A Practical Approach*, Pusat Penerbitan Universiti Teknologi MARA, Shah Alam, Malaysia.
6. Ellis CA, Parbery SA (2005) Is smarter better? A comparison of adaptive, and simple moving average trading strategies. *Res Int Bus Finance* **19**, 399–411.
7. Bowerman BL, O' Connell RT, Koehler AB (2005) *Forecasting, Time Series, and Regression*, 4th edn, Thomson Brooks, Cole, Belmont.
8. Unit Kawalan Penyakit Berjangkit (2006) *Laporan Tahunan TB 2006*, Pejabat Kesihatan Daerah Kuala Terengganu, Kuala Terengganu, Terengganu.