Iranian J Publ Health, Vol. 41, No.10, Oct 2012, pp. 87-96

Original Article

Tuberculosis Surveillance Using a Hidden Markov Model

A Rafei [1], ***E Pasha*** [2], ****R Jamshidi Orak*** [1]

1. *Dept. of Mathematics and Statistics, School of Health Management and Information Sciences, Tehran University of Medical Sciences, Tehran, Iran*
2. *Dept. of Mathematics, School of Mathematics and Computer, Kharazmi University, Tehran, Iran*

***Corresponding Author:** Tel: +98 88 79 38 05 Email: r-jamshidi@tums.ac.ir

(Received 27 Feb 2012; accepted 13 Jul 2012)

Abstract

Background: Routinely collected data from tuberculosis surveillance system can be used to investigate and monitor the irregularities and abrupt changes of the disease incidence. We aimed at using a Hidden Markov Model in order to detect the abnormal states of pulmonary tuberculosis in Iran.

Methods: Data for this study were the weekly number of newly diagnosed cases with sputum smear-positive pulmonary tuberculosis reported between April 2005 and March 2011 throughout Iran. In order to detect the unusual states of the disease, two Hidden Markov Models were applied to the data with and without seasonal trends as baselines. Consequently, the best model was selected and compared with the results of Serfling epidemic threshold which is typically used in the surveillance of infectious diseases.

Results: Both adjusted R-squared and Bayesian Information Criterion (BIC) reflected better goodness-of-fit for the model with seasonal trends (0.72 and -1336.66, respectively) than the model without seasonality (0.56 and -1386.75). Moreover, according to the Serfling epidemic threshold, higher values of sensitivity and specificity suggest a higher validity for the seasonal model (0.87 and 0.94, respectively) than model without seasonality (0.73 and 0.68, respectively).

Conclusion: A two-state Hidden Markov Model along with a seasonal trend as a function of the model parameters provides an effective warning system for the surveillance of tuberculosis.

Keywords: Sputum, Pulmonary tuberculosis, Hidden Markov model, Cyclic regression, EM-algorithm

Introduction

The necessity of tuberculosis (TB) control as one of the top ten mortality causes is completely apparent (1-3). Despite the dramatic advancements in case detection and effective treatment methods against TB over the past six decades (4, 5), it still remains as a great area of concern resulting from the emergence of both multidrug-resistant (MDR) (6-8) and HIV-associated TB (9, 10) in many areas of the world (11). It is estimated about one-third of the people carry the TB bacilli, with almost nine million new infections and two million deaths every year due to TB (1). Moreover, 2.5% of the global burden of disease is attributed to TB (12). In Iran, although implementing National Tuberculosis Program (NTP) (13) caused TB currently to be well-controlled and the associated mortality exhibits a downward trend during past two decades (14), it continues to be a serious menace to public health (15, 16). Increasing the number of MDR-TB (17, 18) and bordering countries with high burden of the disease like Afghanistan, Pakistan and Iraq create particular challenges for the disease control (19, 20). In 2010, a total of 11,092 old and new cases of TB were re-

ported in Iran and of these cases 246 patients (around 2.2%) were HIV positive (21).

TB surveillance and preventing further spread of the disease requires full understanding of the biological factors affecting TB, and also finding mathematical patterns explaining the mechanism of TB transmission through the community (22). Although TB is not recognized as an infection with rapid dynamics (the chance of getting infection per contact is low), the risk of transmission is higher from patients with sputum smear–positive (SS+) (23). Therefore, regarding this assumption, it would be acceptable the fact that the number of persons getting infection over the next time period depends on the number of infectious cases at the current time period. Moreover, some studies have illustrated variable periods of peak seasonality of TB time series with various patterns in different countries. In particular, some of them have reported a higher incidence of the disease in the late winter to early spring in the sense that the indoor activities in the cold weather is much more common than in a warm climate (24, 25).

The early detection of disease outbreaks has been one of the major concerns of all surveillance systems. Since two decades ago, traditional surveillance techniques were replaced by biosurveillance system with the purpose of reducing time delay to detect and report outbreaks (26). Biosurveillance provide early warning system of epidemics by monitoring the data typically consist of time series counts of incident cases of disease, gathered monthly, weekly, or more frequently (27). There has been already a surge of interest and research in using statistical methods for the early detection of outbreaks based on the routinely surveillance data. Regression techniques, time series analysis, statistical process control and Bayesian methods are some examples of the statistical method which have been used for monitoring the epidemiologic surveillance of infectious diseases (28).

Whereas infectious diseases mostly lie into one of the two non-epidemic and epidemic phases (29), it seems using the concept of finite mixture model is preferred to fit models based on a unique distribution. The basic idea of using Hidden Markov Model (HMM) for monitoring the epidemiologic surveillance of infectious diseases was proposed by Le Strat and Carret in 1999 (30). They applied the model to the time series of flu-like disease incidence rates and poliomyelitis counts and demonstrated the ability of HMM in modeling the routinely diseases surveillance data. Nevertheless, it has been rarely applied in public health systems for the same purpose (31). Five years later, Toni M. Rath et al. indicated some problems and shortcomings in Strat's approach and presented some modification to their models (32).

In this study, HMMs which seems to be an appropriate tool in this issue were used to monitor the anomaly states of the weekly numbers of newly diagnosed cases with SS+.

Material and Methods

Study area and data source

The data obtained from the national TB surveillance program of Iran including the entire SS+ patients registered at all TB register offices throughout Iran. The diagnostic criterion of a new SS+ pulmonary TB case was based on the existence of one of the following conditions: (a) At least two initial sputum smear tests positive for Acid-Fast-Bacilli (AFB), or (b) One sputum smear test positive for AFB plus radiographic diagnosis of active pulmonary TB, or (c) One sputum smear positive for AFB plus sputum culture-positive for M. TB. Currently, the information of all new diagnosed cases with any type of TB is recorded and documented at TB register offices located in all counties of Iran and send as a report quarterly to the "Administration of Tuberculosis and Leprosy Control" of Iran Ministry of Health and Medical Education (13).

The time series of the weekly number of newly diagnosed SS+ patients were extracted from April 2005 to March 2011 using the *TB* Register software (version 7). The first version of the software was released in 2005 in order to improve the quality of data and statistical reports resulting from the TB surveillance system of Iran. Fig. 1 exhibits the time series of the weekly number of newly diag-

nosed SS+ patients during the six-year period 2005-2011.

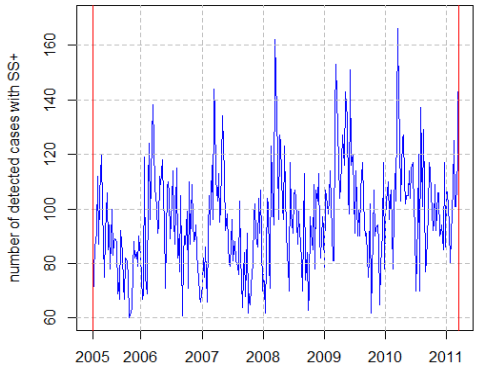

Fig. 1: Weekly number of patients newly diagnosed with SS+ pulmonary TB in Iran over 2005-2011

Cyclic regression

In the mid-1960s, Surfling developed a cyclic regression model to monitor and detect the anomaly activities of Pneumonia and Influenza based on the weekly excess mortality data attributed to the disease. The model characterized the historical sequence of the disease time series by combination of a linear term with a trigonometric function describing the seasonal trend (33). This idea originated from the Fourier series and can be formulated with two terms as a multiple linear regression:

$$y_t = \beta_0 + \beta_1 t + \beta_2 \sin(\frac{2\pi t}{r}) + \beta_3 \cos(\frac{2\pi t}{r}) + \varepsilon_t$$

Where y_t denotes observed weekly Pneumonia and Influenza deaths at week t for five years period; β_j; $j=0,1,2,3$ are regression coefficients, as β_0 and β_1 describe the linear part and β_2 and β_3 belong to the seasonal part; r is the time duration of fluctuations; and ε_t is an independent normally distributed error term with a constant variance.

In fact, the Serfling method followed a two-step procedure. The first step was to determine a baseline describing the expected pattern of the historical disease excess mortality. Since the baseline model estimated the non-epidemic phase of the disease, weeks that typically showed high disease incidence were excluded to avoid overestimating the parameters. The main problem in Serfling approach was to determine the epidemic points in this stage. As a criterion, Pelat et al. (2007) proposed excluding the 20% highest values of data to account for past outbreaks in modeling the baseline (34). Then the estimated baseline was used to predict future time series of the expected disease rates. In the second step, an epidemic threshold was obtained by calculating an upper percentile for the prediction distribution according to the baseline. Consequently, an outbreak was detected while an observation exceeds the predefined threshold (28). Moreover, an automated web-based application of Serfling model has been constructed by Pelat et al. for the retrospective and prospective surveillance of diseases (34).

Hidden Markov Models

A HMM is a statistical tool to fit a mixture distribution on a sequence of dependent data. The application of the models have been recognized in many areas, including automatic speech recognition, electrocardiographic signal analysis, epileptic seizure frequency analysis, DNA sequence analysis, the modeling of neuron firing and meteorology (30). A HMM consists of a bivariate discrete time process like $\{S_t, Y_t\}_{t\geq 1}$, where $\{S_t\}$ is an unobservable Markov chain and, conditional on $\{S_t\}$, $\{Y_t\}$ is a sequence of independent random variables such that the conditional distribution of Y_t only depends on S_t. The sequences $\{S_t\}$ and $\{Y_t\}$ are often called state sequence and observed sequence, respectively (35). The dependence structure of a HMM can be represented by a graphical model as in Fig. 2.

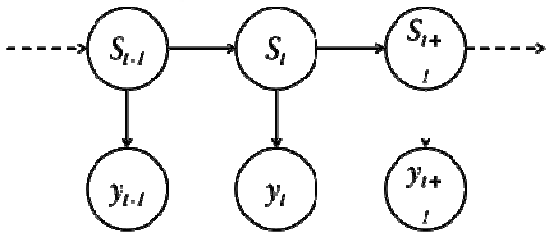

Fig. 2 : Graphical representation of the dependence structure of two sequences in a hidden Markov

Let S_t $(t=1,2,\ldots,n)$ represents a first order Markov chain which takes on one of the m values 1,2,…,m by a transition matrix $\Gamma=(a_{ij})_{m*m}$ and initial probability distribution $\pi=(\pi_1,\ldots,\pi_m)^T$, where

$a_{ij} = P(S_t=j \mid S_{t-1}=i)$ $i,j = 1,2,\ldots,m$; $t = 1,2,\ldots,n$

$\pi_i = P(S_1=i)$ $i = 1,2,\ldots,m$

Moreover, the conditional distribution of Y_t given $S_t=i$ follows a parametric form $f_i(y_t;\theta_i)$ where θ_i is a vector of unknown parameters. Fitting a HMM to the data requires estimation of the parameters including the initial and transition probabilities and distribution parameters. There are mainly two approaches to estimate the parameters in the HMM literatures. The first can be achieved by a maximum likelihood technique using a modified EM-algorithm, known as the Baum-Welch method. The other is Bayesian framework which assumes the parameters follow a prior distribution and then updates them through a Monte Carlo Markov Chain (MCMC) technique. Consequently, after the estimation of parameters, the most likely sequence of hidden states that produced the data should be decoded by use of the Viterbi algorithm. The following section explains how we used HMMs to detect unusual states of sputum smear positive pulmonary TB in Iran.

Fitted model

Two different occurrences were assumed for the disease in every week: *usual phase* which corresponds to the expected disease incidence and *unusual phase* which occurs when the disease incidence increases significantly compared to the usual phase. we described $\{Y_t\}$ at each time $t=1,2,\ldots,n$ to represent the observed number of new SS+ infections occurred over the week t, as Y_t belongs to two distinct distributions associated with the disease states which have been mixed together. A dichotomous variable $\{S_t\}$ was also defined to denote the usual and unusual phases of disease by taking two values 1 and 2, respectively. We accepted $\{S_t\}$ follows a Markov process, on condition that the disease phase in each week depends only on the previous state. In fact, the state sequence of disease $\{S_t\}$ is a virtual variable and cannot be measured directly, because there is no specified criterion for TB to decide on either a usual or unusual state occurs based on the observed number of cases. Again, we described $\pi=(\pi_1,\pi_2)^T$ to be the vector of initial probabilities, where π_1 and π_2 exhibit the probabilities of arising usual and unusual states in the first week, respectively. And $\Gamma=(a_{ij})_{2*2}$ denotes the transition probabilities matrix, where a_{ij} ; $i,j=1,2$ is the transition probability from state i to state j. Consequently, a Hidden Markov Model was obtained by assuming the conditional variables $Y_1, Y_2, \ldots, Y_n$, given the disease state S_t, are independent random variables. We aimed to make an inference about the disease state at each time according to the observed disease incidence. Fig. 3 illustrates a better representation for the first chain of the model structure and related probabilities applied to the SS+ pulmonary TB data.

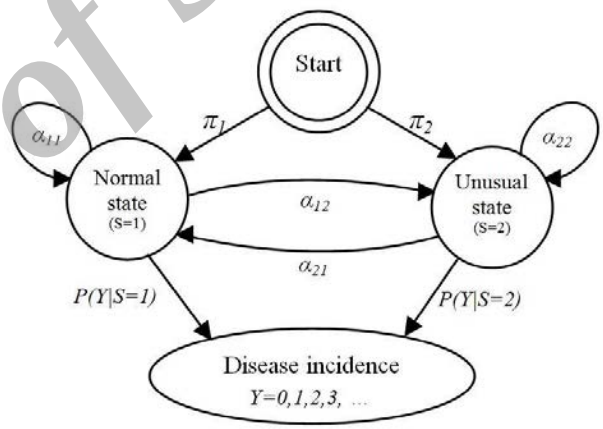

Fig. 3 : Graphical representation of the first chain of HMM structure and related probabilities applied to the SS+ pulmonary TB data

Since the data were considered as discrete counts, a mixture of Poisson-Poisson distribution was assumed to fit to the data. Thus, with a Poisson distribution model we have

$$P(Y_t = s \mid S_t = i) = e^{-\lambda_i}\lambda_i^s/s! \qquad i=1,2$$

In order to consider the fluctuations of SS+ counts time series, we defined two baseline models for evaluating the expected disease counts (λ_t) and then, a Hidden Markov Model was applied for each desired model. The first model was obtained by assuming the expected disease load is

constant over the considered time period; that is the time series of SS+ counts does not follows a seasonal trend. The model parameters in this case, are λ_1 and λ_2 that should be estimated simultaneously along with the related initial and transition probabilities. For the second model, we supposed the baseline of the number of new infections follows a seasonal trend over the considered time period. Since the model parameters were considered time-dependent in this case, another index (t) was added to the model parameters describing the time. We assumed then, a cyclic regression as a function of the model parameters for both states of the disease. Thus, for each state, the model can be written as follows:

$$\lambda_{1t} = E(Y_t \mid S_t = 1) = \beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi t}{r}\right) + \beta_3 \cos\left(\frac{2\pi t}{r}\right)$$

$$\lambda_{2t} = E(Y_t \mid S_t = 2) = (\beta_0 + \beta_e) + \beta_1 t + \beta_2 \sin\left(\frac{2\pi t}{r}\right) + \beta_3 \cos\left(\frac{2\pi t}{r}\right)$$

Where β_i; $i=0,...,3$ are the regression coefficients and β_e reflects an increase in mean upon transition to an unusual phase. After constructing the two baselines, maximum likelihood estimation of all parameters were computed for both models according to the observed sequence of disease incidence using a Baum-Welch algorithm. Besides, in order to detect the anomaly phases of the disease, the most likely sequence of hidden states was detected by Viterbi algorithm based on the updated estimation of parameters.

In addition, in order to evaluate and compare these two models, we measured the models goodness-of-fit using two criteria. The first measure we used was adjusted coefficient of determination. It can be calculated as follows:

$$Adjusted\ R^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$$

Where R^2 denotes the coefficient of determination; n is the number of time points; and k stands for the number of free parameters in the model. The Bayesian Information Criterion (BIC) used as the second criterion. If l_{max} denotes the maximized likelihood and k the number of free parameters, so BIC can be formulated as below:

$$BIC = l_{\max} - \frac{\log(n)k}{2}$$

Then the model with the highest values of both adjusted R^2 and BIC was chosen as the appropriate one. Finally, the results of the selected model were compared with the Serfling epidemic threshold and both sensitivity and specificity indices were computed to check the model validity.

All computations were performed using R version 2.14.1 (Free GNU license) and the source code is also available on request.

Results

The time series of SS+ data included 312 weeks associated with the six years of consideration from April 2005 to March 2011. Fig. 1 illustrates this time series from 2005 to 2011. According to the results, the observed weekly numbers of patients in Iran ranged between 60 and 168 cases with a mean and standard deviation of 97 and 20 cases per week. Table 1 show the monthly number of new SS+ cases and associated incidence rates during six-year period 2005-2011. The population estimations of Iran were obtained from the two nationwide censuses 1996 and 2006 for each year.

Table 1: the number of new SS+ cases and associated incidence rates for each of the six periods 2005-2011

Period*	Number of SS+ cases	Population size	Incidence rates per 100,000
2005-2006	4,561	69,390,405	6.57
2006-2007	4,811	70,495,782	6.82
2007-2008	4,677	71,532,062	6.54
2008-2009	4,880	72,583,586	6.72
2009-2010	5,086	73,650,566	6.91
2010-2011	5,171	74,733,230	6.92

* Each period starts from the beginning of April to the end of the next March.

Two HMMs were applied to the data by assuming both constant and seasonal trend as the expected case load of the disease. The updated estimation of parameters were obtained using Baum-Welch algorithm after 50 and 152 iterations, respectively. The results of both models have been summarized in Table 2.

Table 2: Parameters estimation of two models with and without seasonality for both phases of the disease

Model	Disease phase	Number of states	Initial probably	Transition probability		Model parameters			
				Usual	Unusual	β_0	β_1	β_2	β_3
With seasonality	Usual	192	1	0.70	0.30	76.82	0.07	7.15	7.42
	Unusual	120	0	0.45	0.55	100.26	0.07	7.15	7.42
						λ			
Without seasonality	Usual	160	1	0.79	0.21	82.45			
	Unusual	152	0	0.21	0.79	111.24			

Moreover, the shifts in the mean number of patients while moving to the abnormal phase were observed 23.44 and 28.79 in two models with and without seasonality assumption, respectively. Furthermore, in order to reveal the unusual phases of SS+ pulmonary TB, the most likely state sequence of models was detected by using the updated parameters in Viterbi algorithm. Fig. 4 and 5 present the disease states, detected according to constant and seasonal baselines, respectively.

Consequently, both adjusted R-squared and BIC were measured for the evaluation of the models goodness-of-fit. As Table 3 exhibits, the higher values of both criteria for the model with seasonal trend show a better goodness-of-fit than the model without seasonality.

Table 3: Model evaluation and goodness-of-fit for both models with and without seasonal trends as baseline

Model	Number of iterations	R-squared	Adjusted R-squared	Log-likelihood	BIC
With seasonality	577	0.72	0.72	-1316.56	-1336.66
Without seasonality	50	0.56	0.55	-1375.26	-1386.75

Therefore, the seasonal model was selected to compare with the Serfling epidemic threshold in the detection of disease anomaly states. Also, the upper 95th percentile of the prediction distribution from the seasonal baseline in the HMM was calculated as the Serfling threshold. Considering the threshold results as reference, Sensitivity and specificity were obtained 0.94 and 0.87, respectively. More information has been detailed in Table 4.

In order to achieve a better illustration of variability and fluctuations of SS+ pulmonary TB incidence during a year in Iran, we computed the average for the number of weeks being in unusual phase for each month over 2005-2011.

Table 4: Model validity indices for both baselines based on the Serfling threshold as reference

Model	Validity Index	
	Sensitivity	Specificity
With seasonality	0.94	0.87
Without seasonality	0.68	0.73

Fig. 6 was drawn to depict the changes of the proportion of weeks with unusual phases over months. For both models, the highest average for the number of unusual phases of the disease was obtained in March.

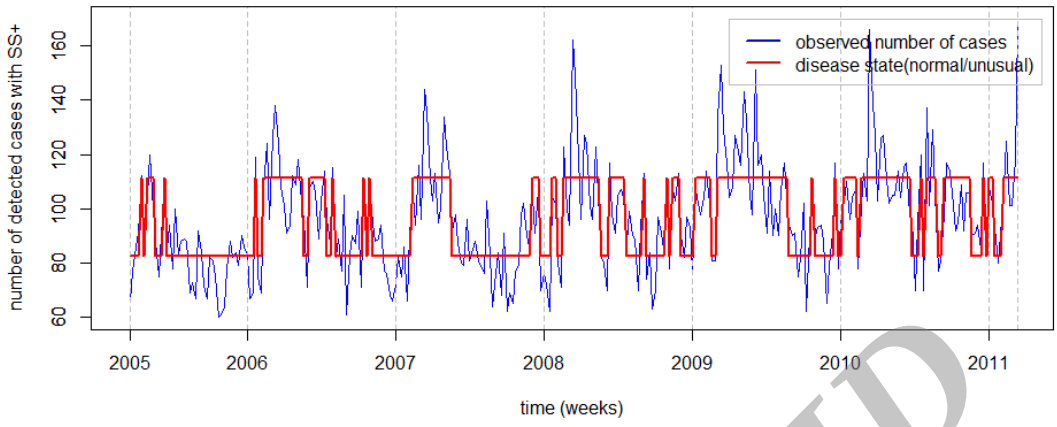

Fig. 4: Unusual states of the weekly SS+ data in Iran over 2005-2011, obtained by applying by Viterbi algorithm in HMM without seasonality

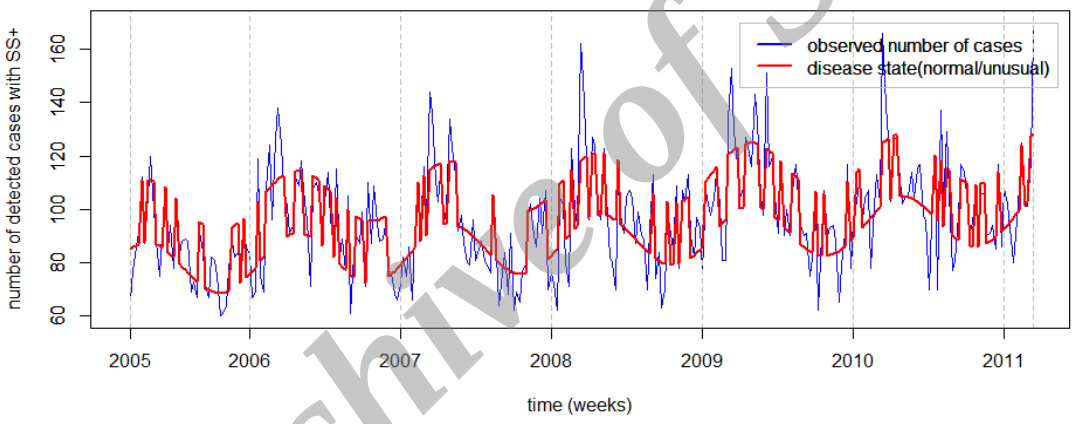

Fig. 5: Unusual states of the weekly SS+ data in Iran over 2005-2011, obtained by applying by Viterbi algorithm in HMM with seasonality

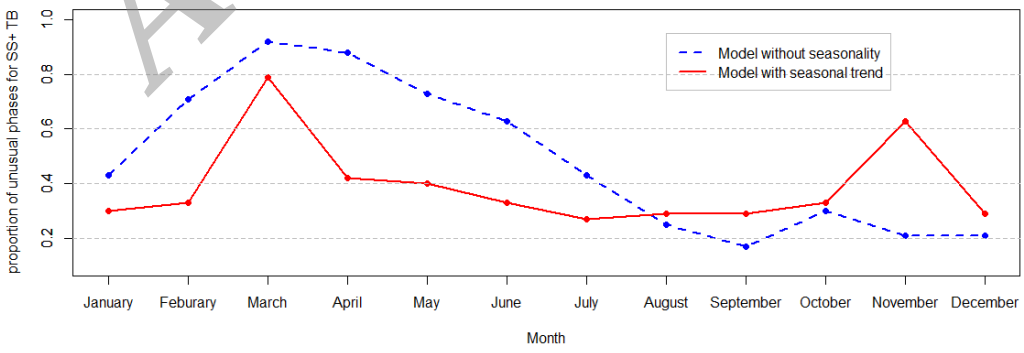

Fig. 6: Comparison of the proportion of unusual phases attributed to SS+ pulmonary TB among months for both baselines, Iran, 2005-2011

Discussion

In this paper, a powerful statistical tool was employed to monitor the routinely TB surveillance data retrospectively, in an attempt at detecting abnormalities and aberrations from the expected disease incidence. Although HMMs have been already applied to the flu-like disease, poliomyelitis (30), Malaria, Leprosy (36), nosocomial infections (37) and Hepatitis A (31) data, we found no literature using the model in TB surveillance for the same purpose. This study demonstrated that a two-state HMM provides a conceptually simple framework to segment the sequence of disease counts of SS+ patients into expected and abnormal phases.

In order to consider the natural variations of TB in Iran, Two models with and without seasonal trends were applied and parameters were trained using the data. The comparison of two models showed a better goodness-of-fit for the model which characterizes baseline with a seasonal trend, having an adjusted R-squared of 0.72 and a BIC of -1336.66 (Table 3). As Fig. 5 shows, the SS+ counts data followed a seasonal trend during this six-year period 2005-2011. The trend characterized the disease time series with a considerable increase in the early of spring. A study from the South Africa also showed seasonal variation of TB rates with a higher incidence in the late winter to early of spring (38). The findings also presented a linear increase in the number of SS+ patients with approximately 7 cases per 100 weeks over the time period of the current study.

According to Fig. 6, we found different patterns for two models in detection of unusual phases of disease. Under the constant assumption for baseline, the model demonstrates an upward trend in the number of detected cases with unusual phases from January and reaches its peak in March and then, decreases gradually by the end of the year. In the model with seasonality assumption, the average for the number of detected weeks with unusual phases exhibits a stable situation over months, except two considerable peaks in March and November. On average, the number of unusual states detected by this model is lower than one without seasonal trends, especially during January to August. Nevertheless, both models we applied in this study show the same peak in March in the average of the weeks with unusual states. Therefore, we may conclude again that a higher incidence in SS+ pulmonary TB is occurred in the late winter in Iran.

Since the data we used for the study presented the weekly number of infections, so a mixture of Poisson probability distribution was assumed to be followed by data. However, the large amounts of the weekly disease counts have caused a lack in the Model goodness-of-fit and made the likelihood to show a very small value. If we could limit our study population to a smaller region like a province or prisoner population, which are at higher risk of TB, we would expect to obtain a better goodness-of-fit. Another way was to use the disease incidence rates rather than disease counts. In this case, there were more available distributions like Gaussian, exponential or a mixture of them to fit on the data. Another attractive alternative was to add other variables such as autoregressive terms or some covariates closely associated with TB like socioeconomic status or climate changes as a function of the model parameters to present better fitness for the disease baseline.

In this study, we used a modified EM-algorithm known as Baum-Welch in some HMM literatures to estimate the model parameters based on a maximum likelihood technique. As a limitation of the study, we did not compute the standard errors of the estimations in order to make inference or report their confidence intervals. Because it needs some other extensions like using bootstrap method and is not addressed in the present study (37). Another alternative is using Bayesian approach which provides simpler framework for the estimation of parameters and making inference about them by a MCMC sampling algorithm (31, 36, 39, 40). To this end, first a prior distribution should be introduced to the parameters and then updated by a MCMC algorithm given the observations.

Ethical considerations

Ethical issues (Including plagiarism, Informed Consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc) have been completely observed by the authors.

Acknowledgement

This study was a part of M.S. thesis supported by Tehran University of Medical Sciences (grant No: 625). We would like to thank Dr. Mohammad Mehdi Gouya and Dr. Saeid Sharafi from the Iran Center for Communicable Diseases control and prevention for providing us with the data. We are also grateful to Dr. Seyed Mohammad Alavinia for reviewing the manuscript and his useful advices as well as Dr. Fereshteh Asgari from the Center for Non-Communicable Diseases Control and Prevention. The authors declare that there is no conflict of interests.

References

1. World Health Organization (2011). Global tuberculosis control: WHO report 2011. *Sixteenth annual report on TB control*, WHO.
2. Mathers C, Fat DM, Boerma J (2008). The global burden of disease: 2004 update. WHO, 8-11.
3. Pinheiro P, Mathers CD, Krämer A (2010). The Global Burden of Infectious Diseases. *Modern Infectious Disease Epidemiology,* pp. 8-19.
4. Migliori G, Loddenkemper R, Blasi F, Raviglione M (2007). 125 years after Robert Koch's discovery of the tubercle bacillus: the new XDR-TB threat. Is "science" enough to tackle the epidemic? *ERJ,* 29(3): 423-7.
5. Daniel TM (2006). The history of tuberculosis. *Respiratory Medicine*, 100(11): 1862-70.
6. Lönnroth K, Castro KG, Chakaya JM, Chauhan LS, Floyd K, Glaziou P, et al. (2010). Tuberculosis control and elimination 2010–50: cure, care, and social development. *The Lancet*, 375(9728): 1814-29.
7. World Health Organization (2010). Multidrug and extensively drug-resistant TB(M/XDR-TB). WHO.
8. Zignol M, Hosseini MS, Wright A, Lambregts–van Weezenbeek C, Nunn P, Watt CJ, et al. (2006). Global incidence of multidrug-resistant tuberculosis. *JID*, 194(4): 479-85.
9. Karim SSA, Churchyard GJ, Karim QA, Lawn SD (2009). HIV infection and tuberculosis in South Africa: an urgent need to escalate the public health response. *The Lancet*, 374(9693): 921-33.
10. Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, et al. (2003). The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. *Arch Intern Med*, 163(9): 1009.
11. Dye C, Lönnroth K, Jaramillo E, Williams B, Raviglione M (2009). Trends in tuberculosis incidence and their determinants in 134 countries. *Bull World Health Organ*, 87(9): 683-91.
12. Montoro E, Rodriguez R (2007). Global burden of tuberculosis. In *Tuberculosis 2007, from basic science to patient care*. Edited by Palomino JC, Leão SC, Ritacco V. Bernd Sebastian Kamps and Patricia Bourcillier, p. 263.
13. Nasehi M, Mirhaghani L (2009). *National guidelines for TB control.* Iran Ministry of Health, pp. 19-20. (*in persian*)
14. World Health Organization (2012). Tuberculosis profile of Iran (Islamic Republic of). WHO.
15. Velayati AA, Masjedi MR, Farnia P, Tabarsi P, Ghanavi J, ZiaZarifi AH, et al. (2009). Emergence of new forms of totally drug-resistant tuberculosis bacilli. *Chest*, 136(2): 420-5.
16. Velayati A, Farnia P, Masjedi M, Ibrahim T, Tabarsi P, Haroun R, et al. (2009). Totally drug-resistant tuberculosis strains: evidence of adaptation at the cellular level. *ERJ*, 34(5): 1202-3.
17. AR HT, Yari S, Karimi A, Fateh A, Bahrmand A, Saifi M, et al. (2011). Survey of extensively drug-resistant tuberculosis (XDR-TB) in Iran-Tehran: A retrospective study. *Afr J Microbiol Res*, 5(22): 3795-800.
18. Mirsaeidi MS, Tabarsi P, Farnia P, Ebrahimi G, Morris MW, Masjedi MR, et al. (2007).

Trends of drug resistant Mycobacterium tuberculosis in a tertiary tuberculosis center in Iran. *Saudi Med J*, 28(4): 544.
19. Masjedi MR, Farnia P, Sorooch S, Pooramiri MV, Mansoori SD, Zarifi AZ, et al. (2006). Extensively drug-resistant tuberculosis: 2 years of surveillance in Iran. *Clin Infect Dis*, 43(7): 841.
20. Khazaei HA, Rezaei N, Bagheri GR, Dankoub MA, Shahryari K, Tahai A, et al. (2005). Epidemiology of tuberculosis in the southeastern Iran. *Eur Journal Epidemiol*, 20(10): 879-83.
21. Iran Center for Communicable Diseases Control (2010). The status of TB/HIV Co-infection. Iran Ministry, Available from: http://www.cdc.hbi.ir. (*in persian*)
22. Aparicio JP, Castillo-Chavez C (2009). Mathematical modelling of tuberculosis epidemics. *Math Biosci Eng*, 6: 209-37.
23. Hertzberg G (1957). The infectiousness of human tuberculosis; an epidemiological investigation. *Acta Tuberculosea Scandinavica Supplementum*, 38:1.
24. Liu L, Zhao XQ, Zhou Y (2010). A tuberculosis model with seasonality. *Bull Math Biol,* 72(4): 931-52.
25. Rios M, Garcia J, Sanchez J, Perez D (2000). A statistical analysis of the seasonality in pulmonary tuberculosis. *Eur J Epidemiol*, 16(5): 483-8.
26. Shmueli G, Burkom H (2010). Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics*, 52(1): 39-51.
27. Castillo-Chavez C (2010). *Infectious Disease Informatics and Biosurveillance*. Springer Verlag, pp. 3-8.
28. Unkel S, Farrington C, Garthwaite PH, Robertson C, Andrews N (2011). Statistical methods for the prospective detection of infectious disease outbreaks: a review. *J Roy Statist Soc Ser A*, 175(1): 49-82.
29. Lu HM, Zeng D, Chen H (2009). Prospective infectious disease outbreak detection using Markov switching models. *IEEE T Knowl Data En*, 565-77.
30. Le Strat Y, Carrat F (1999). Monitoring epidemiologic surveillance data using hidden Markov models. *Stat Med*, 18(24): 3463-78.
31. Watkins R, Eagleson S, Veenendaal B, Wright G, Plant A (2009). Disease surveillance using a hidden Markov model. *BMC Med Inform Decis Mak*, 9(1): 39.
32. Rath T, Carreras M, Sebastiani P (2003). Automated detection of influenza epidemics with hidden Markov models. *CHES*, 521-32.
33. Serfling RE (1963). Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public health reports,* 78(6): 494.
34. Pelat C, Boëlle PY, Cowling B, Carrat F, Flahault A, Ansart S, et al. (2007). Online detection and quantification of epidemics. *BMC Med Inform Decis Mak*, 7(1): 29.
35. Cappé O, Moulines E, Rydén T (2005). *Inference in hidden Markov models*. Springer Verlag, 1-3.
36. Jamshidi Orak R, Mohammad K, Pasha E, Sun W, Nori Jalyani K, Rasolinejad M, et al. (2007). Modeling the spread of infectious diseases based the Bayesian approach. *J School Public Health Research*, 5(1): 7-15. (*in persian*)
37. Cooper B, Lipsitch M (2004). The analysis of hospital infection data using hidden Markov models. *Biostatistics*, 5(2): 223-37.
38. Schaaf H, Nel E, Beyers N, Gie R, Scott F, Donald P (1996). A decade of experience with Mycobacterium tuberculosis culture from children: a seasonal influence on incidence of childhood tuberculosis. *Tubercle and Lung Disease*, 77(1): 43-6.
39. Martínez Beneito MA, Conesa D, López Quílez A, López Maside A (2008). Bayesian Markov switching models for the early detection of influenza epidemics. *Stat Med*, 27(22): 4455-68.
40. Held L, Hofmann M, Höhle M, Schmid V (2006). A two-component model for counts of infectious diseases. *Biostatistics*, 7(3): 422-37.