Infection, Disease & Health (2016) **21**, 184–191

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: http://www.journals.elsevier.com/infection-disease-and-health/

Research

Parameter estimation of tuberculosis transmission model using Ensemble Kalman filter across Indian states and union territories

Pankaj Narula [a], **Vihari Piratla** [b], **Ankit Bansal** [c], **Sarita Azad** [a,*], **Pietro Lio** [d]

[a] *School of Basic Sciences, Indian Institute of Technology Mandi, Mandi 175001, Himachal Pradesh, India*
[b] *School of Computing and Electrical Engineering, Indian Institute of Technology Mandi, Mandi 175001, Himachal Pradesh, India*
[c] *Department of Mechanical and Industrial Engineering, Indian Institute of Technology Roorkee, Roorkee 247667, Uttarakhand, India*
[d] *Computer Laboratory, William Gates Building, 15 JJ Thomson Avenue, Cambridge CB3 0FD, University of Cambridge, UK*

Received 6 August 2016; received in revised form 6 November 2016; accepted 7 November 2016
Available online 7 December 2016

KEYWORDS
Tuberculosis;
India;
Infection

Abstract *Background:* Tuberculosis (TB) is one of the main causes of mortality on the globe. Besides the full implementation of Revised National Tuberculosis Control Programme (RNTCP), TB continues to be a major public health problem in India.
Methods: In the present study, parameters of a TB model are estimated using Ensemble Kalman filter (EnKf) approach. Infection rate and fraction of smear positive cases of TB are estimated in context of India.
Results and Conclusions: Results reveal that the infection rate is highest in Manipur and the ratio of smear positive cases is highest in Pondicherry. The infection rate of TB in Manipur is found to be 2.57 per quarter for the period 2006–2011.

Highlights

- Deterministic TB model is used to model the TB transmission in India.
- Parameters of model are estimated using Ensemble Kalman filter for the period 2006–2011.
- Infection rate is found to be highest in Manipur.

* Corresponding author.
E-mail address: sarita@iitmandi.ac.in (S. Azad).

http://dx.doi.org/10.1016/j.idh.2016.11.001
2468-0451/

Introduction

Tuberculosis (TB) is a well known infectious disease caused by bacterium *M. tuberculosis* which generally spreads through air. In 2011, 2.0—2.5 million new TB cases were estimated in India out of the global annual incidence of 9.4 million cases [1,2]. A large number of factors contribute to the spread of TB; such as high prevalence of HIV/AIDS and diabetes, poor hygiene, crowding, illiteracy and lack of awareness make the TB situation critical in Indian context. All these factors directly contribute to high infection rates among the population. In 1997, the Government of India, with the help of World Bank, initiated RNTCP based on the internationally recommended *Directly Observed Treatment Short*-course (DOTS) strategy [2,3]. RNTCP is the largest TB control program in terms of treatment of patients with full nationwide coverage.

Mathematical models and statistical techniques play a significant role in understanding the transmission dynamics of TB. In simple deterministic model of infectious disease, the number of susceptible persons who are infected by an infectious individual per unit of time is proportional to the total number of susceptible persons. This proportional coefficient is defined as infection rate. Estimation of parameters of mathematical model, for instance, infection rate contributes to better quantify the spread of disease.

Generally, inference of these parameters is a difficult task because of poor compatibility between observed data and models. Simulations and epidemiological data have been used to estimate the key parameters of deterministic models. Different techniques have been introduced and applied to estimate the parameters of TB models. Approximate Bayesian computation approach has been used to estimate TB transmission rate parameters for United States [4]. A synchronisation based method has been implemented to infer the parameters such as treatment rate, disease induced mortality rate and infection rate of a TB model. In particular, the infection rate in the study is estimated to be 2.04 for the quarterly data during 2003—2007 for Cameroon [5]. Liu et al. estimated the reactivation and infection rate of a TB model for China by assuming these rates as sinusoidal functions and infection rate is estimated to be 2.23 person per month for the period 2005—2009 [6]. A qualitative analysis of the TB model for Nigeria has been performed to analyse the effect of DOTS strategy [7]. Mandal et al. estimated parameters such as infection rate and treatment rate using annual prevalence and incidence data of TB [8]. In particular, the infection rate of TB have been estimated to be 11.03 per year for India [8]. Mishra et al. integrated quarantine compartment into TB model, which incorporates the multidrug-resistant TB patients. The model is further analysed and simulated in using TB data of Jharkhand, India [9].

In the present paper, we use Ensemble Kalman filter (EnKf) approach to estimate the parameters of a deterministic model of TB. Kalman filter has been extensively used to infer the parameters of models of various infectious diseases [10—14]. Parameters of an HIV/AIDS model has been estimated using Kalman filter approach [10]. An extension of Kalman filter has been implemented to analyse the spatio-temporal behaviour of measles outbreak using count data for the period 1960—1970 in Landon [12]. Influenza data of different cities within the United States has been analysed and demonstrated that ensemble filters were found to be more accurate than other filters in predicting the peaks of the influenza [13].

Methods

Dynamic model

In this paper, we use a variation of SIR (Susceptible-Infected-Recovered) model defined as SLIS (Susceptible-Latent-Infected-Susceptible). There are three exclusive groups of individuals; namely, susceptible, S, latently infected, L (infected with *M. tuberculosis* but not infectious), and actively infected with *M. tuberculosis*, I (infected and infectious). The model does not take into account genetic and demographic heterogeneity. The following are the governing differential equations for the rate of change in population in various compartments.

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta SI}{N} + \gamma(I+L) \\
\frac{dL}{dt} &= (1-p)\frac{\beta SI}{N} - \gamma L - \mu L \\
\frac{dI}{dt} &= p\frac{\beta SI}{N} - \gamma I + \mu L
\end{aligned} \quad (1)$$

where β is the transmission rate (number of contacts made by an infectious person per quarter), γ is the recovery rate (assumed to be 0.8 for both active and latent infections) [2], μ is the re-infection rate from latent to active disease (assumed to be 0.1) [15], and N is the total population (assumed to be constant). Patients with the latent form of infection are assumed to develop active tuberculosis at an average rate of t, with a 5—10% lifetime risk of a latent infection reactivating to active TB disease. The rate at which *M. tuberculosis* infected people spread active TB to susceptible people is proportional to p, and the rate at which latent cases occur is proportional to $(1 - p)$. In general, untreated patients can infect 10—15 persons each year [2].

State space formulation

The parameters β and p in the dynamic SLIS model are modelled using state-space formulation and EnKf. These two parameters cannot be observed directly. The methodology employs a two-step forecasting model. The coefficients of the SLIS model are written into a very simple state space model representing a Markov process as

$$\begin{aligned}
a_{t+1} &= \begin{bmatrix} \beta \\ p \end{bmatrix}_{t+1} = \begin{bmatrix} \beta \\ p \end{bmatrix}_t + w_t \\
a_{t+1} &= a_t + w_t
\end{aligned} \quad (2)$$

$$w_t = N\begin{bmatrix} 0.5 \\ 0.02 \end{bmatrix} \quad (3)$$

where w_t is the uncertainty in the model parameters assumed to be given by Gaussian white noise with standard deviations 0.5 and 0.02, respectively. The measurement