Journal of Epidemiology and Global Health (2015) 5, 337–346

http:// www.elsevier.com/locate/jegh

# Analyzing seasonality of tuberculosis across Indian states and union territories

Pankaj Narula [a], Praveer Sihota [a], Sarita Azad [a,*], Pietro Lio [b]

[a] School of Basic Sciences, Indian Institute of Technology Mandi, Mandi 175001, Himachal Pradesh, India
[b] Computer Laboratory, William Gates Building 15, JJ Thomson Avenue, Cambridge CB3 0FD, University of Cambridge, UK

Received 29 November 2014; received in revised form 4 February 2015; accepted 6 February 2015
Available online 18 March 2015

**KEYWORDS**
Tuberculosis;
Seasonality;
Poisson regression;
Smear positive cases;
Temperature

**Abstract** A significant seasonal variation in tuberculosis (TB) is observed in north India during 2006–2011, particularly in states like Himachal Pradesh, Haryana and Rajasthan. To quantify the seasonal variation, we measure average amplitude (peak to trough distance) across seasons in smear positive cases of TB and observe that it is maximum for Himachal Pradesh (40.01%) and minimum for Maharashtra (3.87%). In north India, smear positive cases peak in second quarter (April–June) and reach a trough in fourth quarter (October–December), however low seasonal variation is observed in southern region of the country. The significant correlations as 0.64 (*p*-value < 0.001), 0.54 (*p*-value < 0.01) and 0.42 (*p*-value < 0.05) are observed between minimum temperature and seasonality of TB at lag-1 in north, central and northeast India respectively. However, in south India, this correlation is not significant.


## 1. Introduction

India accounts for 26% of global cases of TB incidences. In 2012, 2.2 million TB cases were estimated in India [1]. Besides the full implementation of Revised National Tuberculosis Program (RNTCP) in 2006, TB still continues to be a leading cause of mortality and economic burden. High prevalence of HIV/AIDS, poverty, diabetes, malnutrition, poor ventilation, lack of awareness and education are some major causes for the vulnerable situation of TB in India.

Seasonal variation in any disease is defined as periodic occurrence in the reported cases with calendar time. TB, in general, is not known as having seasonal patterns like cholera, malaria, measles, chickenpox, rotavirus and diphtheria. However, various studies exhibit the seasonal behavior of TB in different regions of the globe [2,3]. Schaaf et al. [4] analyzed the seasonality of TB in South Africa during the period from 1983 to 1993 and observed high TB incidence in late winter

* Corresponding author. Tel.: +91 1905 237928.
*E-mail address:* sarita@iitmandi.ac.in (S. Azad).

http://dx.doi.org/10.1016/j.jegh.2015.02.004
2210-6006/© 2015 Ministry of Health, Saudi Arabia. Published by Elsevier Ltd.

to early spring. Studies conducted in UK, Hong Kong, Spain and Japan identified summer as peak season of TB notification [5—8]. Mabaera et al. [9] assessed the seasonality of tuberculosis of four countries, namely, Moldova, Mongolia, Uganda and Zimbabwe for the period 1999—2003. Akhtar et al. [10] reported that TB detection peaked in spring season among migrant workers in Kuwait during the period 1997—2006. Cao et al. [11] observed the seasonal pattern in TB data of China for the period 2005—2011 and reported that TB incidence was highest in March—June and lowest in January—February. Li et al. [12] reported the peak to be winter and trough to be autumn season in TB data of China for the period 2005—2012. Willis et al. [13] studied the seasonality of TB of US data for the period 1993—2008.

In the context of India, Thorpe et al. [14] assessed the seasonal patterns in TB for the period 1996—2001 and reported that the behavior of TB is seasonal in northern region of the country with peak occurring in second quarter and trough in fourth quarter of the year; however seasonality was not observed in south India. Recently, Kumar et al. [15] observed a seasonal pattern and declining trend in the TB data of Delhi for the period 2007—2012.

There are several techniques to assess the seasonality in epidemiological data. Liu et al. [16] and Bowong et al. [17] studied the seasonal behavior of TB data with the help of a deterministic model, by taking transmission coefficient of TB as a sinusoidal function. Li et al. [12] and Kumar et al. [15] performed time series analysis to exhibit the seasonal pattern in TB data of China. Cao et al. [11] used a hybrid model, a combination of SARIMA and generalized regression neural network model to examine the seasonality in TB data of China. Christiansen et al. [18] described three widely used techniques to analyze the seasonal effects in the epidemiological data, namely comparison of discrete time periods, geometrical models and generalized linear models (GLMs). Out of these three techniques, GLMs are a class of statistical models that allows studying the seasonality efficiently because of their capability of fitting the data to different underlying mathematical functions [19]. In general, log linear Poisson regression model serves the purpose of underlying function [20]. Christensen et al. demonstrated through simulations that Poisson regression models provide more accurate estimation of seasonal variation as compared to geometrical models [21]. We, therefore, employ Poisson regression model to estimate the seasonal amplitude in TB data of India.

## 2. Method

To estimate seasonal amplitude, we use a generalized linear model with a log link function and Poisson distribution. Poisson regression model is a standard approach to model arrivals in queue. Our data represent the number of incidents of an event over a fixed period of time. We, therefore, use Poisson regression to model the observed data.

We have quarterly data of TB smear positive reported cases represented by $n_t, t = 1, 2, 3, \ldots, 24$ with each $n_t$ corresponding to quarter $t$ has a Poisson distribution with parameter $\lambda_t$.

$$E(n_t) = \lambda_t \quad (1)$$

A Poisson mean denotes the non negative expected count and for this purpose a log-link function is used. To estimate the seasonality in data, we model log of mean of observed data as a sinusoidal function [20]. The Eq. (1) will become

$$\log(\lambda_t) = S_t \quad (2)$$

where $S_t$ is defined to be seasonal component and is given as

$$S_t = \varphi_1 \cos(\theta_t) + \varphi_2 \sin(\theta_t) \qquad \text{for } t = 1, 2, , \ldots, 24 \quad (3)$$

$\theta_t$ is the frequency parameter and $\theta_t = \frac{2\pi t}{K}$, $K$ depends on number of observation in an year. Since we have quarterly reported data, so we choose $K = 4$.

Define $\alpha$ as,

$$\alpha = \sqrt{\varphi_1^2 + \varphi_2^2} \quad (4)$$

The peak to trough difference can be defined as a measure of seasonal variation, which exhibits a comparison between the highest and the lowest value, given as following

$$D = \max_t \exp(S_t) - \min_t \exp(S_t) \quad (5)$$

So from Eqs. (4) and (5), we can estimate peak to trough distance as given below

$$D' = \exp(\alpha) - \exp(-\alpha) \quad (6)$$

## 3. Data

In RNTCP, notification data of TB is received in the form of monthly reports at sub-district level and quarterly reports from sub-district level to district/state level. Paper based reports are used up to district level and from district level, data are transferred electronically to state and center level. We use quarterly reported data of 34 Indian states and union territories (UT's) published by Central TB Division, Directorate General of Health