MODELING THE INCIDENCE OF TUBERCULOSIS IN SOUTHERN THAILAND

Noodchanath Kongchouy[1], Sampurna Kakchapati[2] and Chamnein Choonpradub[2]

[1]Department of Mathematics, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla; [2]Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Pattani, Thailand

Abstract. The aim of this study was to examine the trend, seasonal and geographic effects on tuberculosis (TB) incidence in the fourteen southern provinces of Thailand from 1999 to 2004. Data were obtained from the National Notifiable Disease Surveillance Report (506), Ministry of Public Health. The joint effects of gender, age, quarterly season and location on the TB incidence rates were modeled using both negative binomial distribution for the number of cases and log-linear distribution for the incidence rate; then these models were compared. The linear regression models provided a good fit, as indicated by residual plots and the R^2 (0.64). The model showed that males and females aged less than 25 years had similar risks for TB in the study area. Both sexes had their risk increased with age but to a much greater extent for men than women, with the highest rate noted in males aged 65 years and over. There was no evidence of a trend in the annual incidence of TB during 1999-2004, but the incidence has a significant season variation with peaks in the first quarter over the six year period. There were also differences in the incidence rate of TB both within and between provinces. The high risk areas were in upper western and lower southern parts of the region. The log-linear regression model could be used as a simple method for modeling TB incidence rates. These findings highlight the importance of selectively monitoring geographic location when studying TB incidence patterns.

Key words: log-linear models, negative binomial model, tuberculosis incidence

INTRODUCTION

The scale of the global tuberculosis (TB) epidemic is substantial. About a third of the world's population is infected with *Mycobacterium tuberculosis*, with an estimated nine million new cases and two million deaths occuring yearly due to the disease. It is a leading cause of death among people who are HIV-positive (WHO, 2009b). It affects the mostly economically productive age group comprising adults aged 15 to 54 years, with males being disproportionately affected. The male/female ratio of newly detected cases is 2:1. The impact of TB is particularly evident in Asia (Southeast Asia and the Western Pacific) and Africa (WHO, 2008). Approximately 86% of all TB cases reported worldwide occur in these regions, where 60% of the world's populations live. The Southeast Asia region carries one third of global burden of TB.

Correspondence: Chamnein Choonpradub, Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Mueang 94000, Pattani, Thailand.
Tel: 66 (0) 89466 0803; Fax: 66 (0) 7331 2729
E-mail: nootchanath.k@psu.ac.th, cchamnein@bunga.pn.psu.ac.th

However, TB epidemic in Asia could gradually worsen by variations in the quality and coverage of various TB control interventions, population demographics, urbanization, changes in socio-economic standards, HIV and, more recently, emerging drug resistance (WHO, 2009a).

TB is a serious public health problem in Thailand. The country is ranked 18^{th} on the list of 22 high TB-burden countries. The prevalence of TB was estimated at 192 per 100,000 persons for all forms in 2007, with an incidence rate of 62 new smear-positive cases per 100,000 (WHO, 2009a). Several studies suggest marked regional differences in incidence of TB in Thailand (Wibulpolprasert, 2007). Recent research based on National Tuberculosis Program (NTP) surveillance data from 2001 to 2005 show higher numbers of cases and deaths from TB occurred in the southern and northern provinces of Thailand due to low treatment success rates (Thongraung *et al*, 2008).

The Thailand Health Profile Report indicates that the number of TB cases declined between 1985 and 1989 but increased slightly from 1990 to 2005. The tuberculosis prevalence has risen by 2% each year during the past five years and there was no tendency to decline during the period 1995-2002 (Wibulpolprasert, 2007). Thus, TB remains a major infectious disease in Thailand (Wibulpolprasert, 2007; Phomborphub *et al*, 2008; Thongraung, 2008; Jittimanee *et al*, 2009).

Public health officials are often required to evaluate disease incidence in the country. They compare the standardized disease incidence rate within the area and time frame for planning various interventions. Statistical modeling may provide the necessary quantitative framework for investigating main issues related to disease. Since the 1960s, statistical models have been used to understand tuberculosis transmission dynamics and to predict the effects of different interventions (Waaler *et al*, 1962). The models have been used to understand the impact of tuberculosis control strategies of tuberculosis, (Legrand *et al*, 2008) HIV and TB joint epidemic (Willam *et al*, 2005; Baclear, 2008) and spatial and temporal variation of TB incidence (Nunes, 2007; Uthman, 2008) and drug-resistance tuberculosis (Castillo-Chavez *et al*, 1997; Dye *et al*, 1998).

The purpose of this study was to examine the trend, seasonal and geographic effects on TB incidence in the fourteen southern provinces of Thailand from 1999 to 2004. In our study we used two different statistical models: a negative binomial generalized linear model and a simple linear model after logarithmic transformation of incidence rates, and we compared the results obtained from applying these methods. After fitting the models, we analyzed the findings based on the best fitting model.

MATERIALS AND METHODS

Study area and data source

The study design was a retrospective analysis of reported TB cases obtained from the National Notifiable Disease Surveillance Report (506), Bureau of Epidemiology, Ministry of Public Health, Thailand. The study population included all reported TB cases during the 6-year period 1999-2004 for the 14 southern provinces of Thailand. These data were available as computer files with individual records for disease cases. Fields were comprised of characteristics regarding subject and disease, and included dates of sickness, diagnosis, the subject's age, gender, address, severity of illness and date of death for mortality cases. The resident population

Table 1
Definitions and populations of super-districts.

Code	Super-district	Population
1	Chumphon North	246,279
2	Chumphon South	199,927
3	Ranong	161,210
4	Surat Thani North West	206,713
5	Surat Thani City	241,373
6	Surat Thani East	168,801
7	Surat Thani South	252,523
8	Phang-Nga	234,188
9	Nakhon Si Thammarat North	240,392
10	Nakhon Si Thammarat North West	220,430
11	Nakhon Si Thammarat City	267,560
12	Nakhon Si Thammarat West	273,583
13	Nakhon Si Thammarat South Coast	194,771
14	Nakhon Si Thammarat South West	323,075
15	Krabi	336,210
16	Phuket	249,446
17	Trang South	264,412
18	Trang City	330,698
19	Phatthalung City	251,029
20	Phatthalung West	247,442
21	Songkhla North Coast	329,133
22	Songkhla City	340,096
23	Hat Yai	350,776
24	Songkhla South West Coast	235,657
25	Satun	247,875
26	Pattani City-West	332,757
27	Pattani East	263,228
28	Yala City West	211,180
29	Yala South	204,357
30	Narathiwat Central	224,646
31	Narathiwat West	214,412
32	Narathiwat South	223,292

denominators used to compute incidence rates were obtained from the Population and Housing Census of 2000 undertaken by the National Statistics Office of Thailand. To simplify the effect of the location of residence when calculating incidence rates, smaller contiguous districts in each province were grouped together to form 32 super-districts containing populations ranging from 161,210 to 360,000 (Table 1; listed in order of geographical location from north to south keeping super-districts within the same province).

Age, gender, residential area (by super-district), quarter of the year and year were selected as explanatory variables for studying the incidence rates of TB. Age was divided into four groups (0-24, 25-39, 40-59 and >60 years) and age and gender were combined together to form 8 gender-age groups. The year was divided into four 'quarter' periods, defined as January-March, April-June, July-September and October-December, giving twenty-four quarter periods over the six years.

Statistical methods

We first calculated the disease incidence in cells defined by gender-age group *i*, region *j*, quarterly season *q* and year *t* as the ratio of the number of reported cases n_{ijqt} to P_{ij}, with the corresponding population at risk in 1,000s.

The negative binomial generalized linear model (GLM) (Venables and Ripley, 2002) is an extension of the Poisson Regression Model and allows for over-dispersion. If λ_{ijqt} denotes the mean incidence rate per gender-age group *i*, quarter *q*, super *j* and year *t*, an additive model with this distribution is expressed as:

$$\ln(\lambda_{ijqt}) = \ln(P_{ij}) + \mu + \alpha_i + \beta_q + \gamma_j + \eta_t \qquad (1)$$

The terms α_i, β_q, γ_j and η_t represent gender-age group, season, super-district and year effects, respectively, that sum up to zero, so that μ is a constant encapsulating the overall incidence. The variance of this distribution is $\lambda_{ijqt}(1+\lambda_{ijqt}/\theta)$ with the Poisson model arising in the limit as $\theta \rightarrow \infty$. The model fit was assessed by comparing deviance residuals with normal quantiles, and was informative to plot observed counts and appropriately scaled incidence rates against corresponding fitted values based on the model.

The model also gives adjusted incidence rates for each factor of interest, obtained by suppressing the subscripts in Equation (1) corresponding to the other factors and replacing these terms with a constant satisfying the condition the sum of the disease counts based on the adjusted incidence rates matches the total. Sum contrasts (Venable and Ripley, 2002; Tongkumchum and McNeil, 2009) were used to obtain confidence intervals for comparing the adjusted incidence rates within each factor with the overall incidence rate.

The alternative additive log-linear model for incidence rates with normally distributed errors is:

$$\ln\left(\frac{n_{ijqt}}{P_{ij}}\right) = \gamma_{ijqt} = \mu + \alpha_i + \beta_q + \gamma_j + \eta_t \qquad (3)$$

Since some cells had no reported cases, thus disallowing log-transformation, we replaced the zeros with a suitably chosen small constant *d*, without changing any values for n_{ijqt} greater than 0.

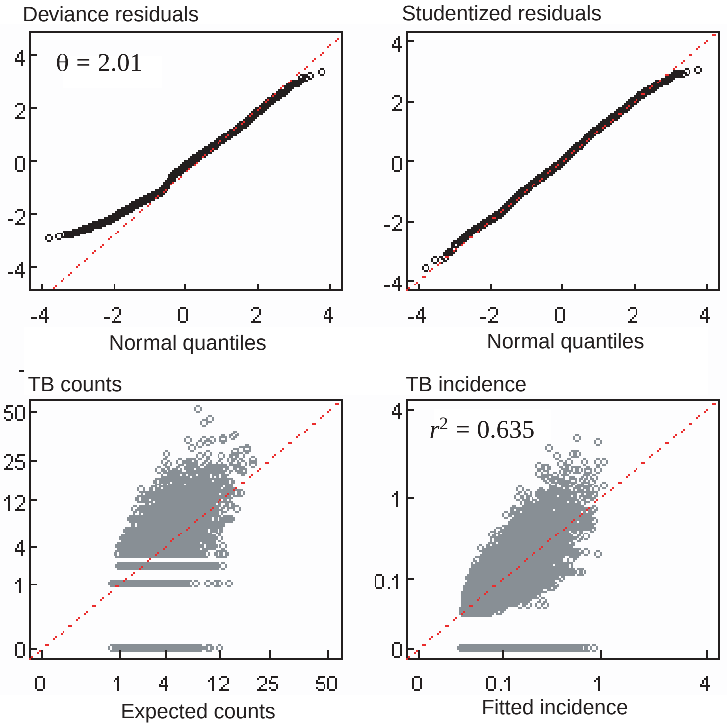
The model fit was assessed by plotting standardized residuals against normal quantiles, and by plotting observed counts and appropriately scaled incidence rates against corresponding fitted values based on the model, and by using r-squared to see how much variation in the data was accounted for by the model.

To obtain estimates of incidence rates in cells we used the formula:

$$\hat{r}_{iqjt} = \exp(\hat{y}_{iqjt} + c), \qquad (4)$$

where y_{ijqt} is the fitted value of y_{ijqt} and *c* is a constant chosen to match the total number of observed cases with the total given by the model.

After fitting the model, incidence rates for levels of each factor adjusting for other factors were calculated similarly. Standard errors for these adjusted incidence rates were obtained using sum contrasts to compare the incidence rates for each level of a factor with the overall mean incidence rate. Using this method, the pattern of TB was identified for each factor. Since the confidence intervals for factor-specific incidence rates obtained in this way divide



Theta (θ) is a dispersion parameter in the negative binomial distribution.

Fig 1–Diagnostic residual plots for negative binomial (top left) and log-normal (top right) models, and plots of counts and incidence rates for the log-normal model (lower panels) for tuberculosis incidence rates in Southern Thailand super-districts.

naturally into three groups according to their location entirely above the mean, around the mean, or entirely below the mean, we used this trichotomy to create thematic maps for districts according to their estimated tuberculosis annual incidence rates.

All models were fitted by maximum likelihood and the analysis was undertaken using purpose-written code in R, version 2.7.1 (R Development Core Team, 2008).

RESULTS

The results of model fitting are shown in Fig 1. The left and right upper panels show plots of deviance residuals against normal quantiles based on the negative binomial model (1) (upper panels) and the log-linear alternative (2) with zero counts replaced by 1 (lower panels). The left and right lower panels show plots of observed counts and observed annual incidence rates per 1,000 population versus corresponding fitted values using the linear model. Clearly, the residuals plot from the negative binomial model did not fit the data as well as the linear model on the log-transformed incidence rates.

Fig 2 shows the 95% confidence intervals for the annual tuberculosis incidence rates by gender-age group (upper left panel), year and season (upper right panel) and super-district (lower panel), each adjusted for the effects of the other two factors in the model based on the log linear model. The dotted horizontal lines on each graph represent the overall mean annual incidence rate (0.45 per 1,000). The dotted vertical lines in the lower panel separate the 14 provinces of southern Thailand. The male to female incidence rate ratio was calculated to be 2.16.

For those aged below 25 years, no gender differences were found. There was a notable difference in incidences between males and females over age 25 years with the highest difference observed in the over age 60 years group (0.87). There was no evidence of a change in trend in the incidence of TB from 1999 to 2004. Higher

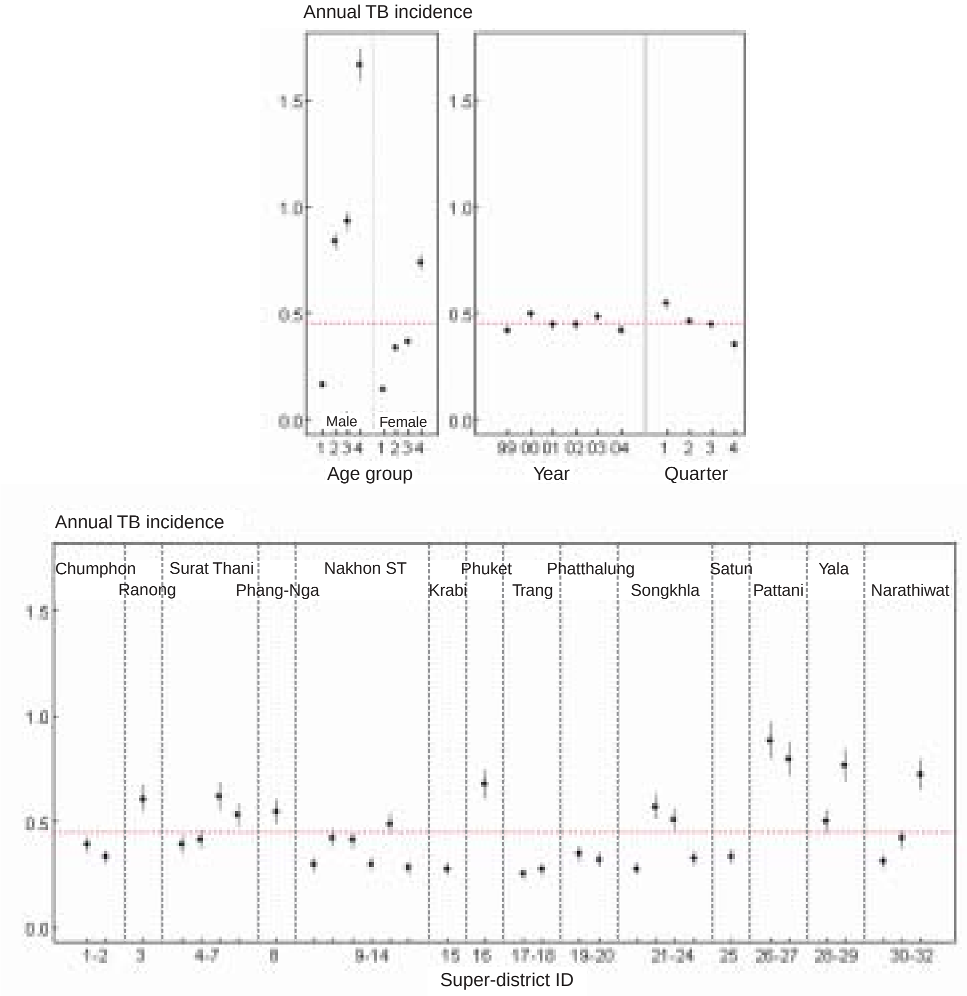


Fig 2–Annual TB incidence/1,000 population in southern Thailand by demographic factor adjusted for other factors, with 95 % confidence intervals for differences from the mean.

incidences occurred during each first quarter, with lower rates during the second half of each year.

Fig 3 shows a thematic map of the adjusted annual incidence by super-district using the confidence intervals plotted in Fig 2 to classify districts as above the mean (darkest shade), below the mean (lightest shade) or near the mean (intermediate shade). Higher TB incidence occurred in the super districts of Pattani, Ranong, Phang-Nga, Phuket and Yala Provinces. Hat Yai and Songkhla City had higher TB incidences. The super-districts of Phang-Nga, Ranong and Phuket Provinces also had higher than average incidence rates.

DISCUSSION

This study used statistical modeling of TB incidence to gain insights into its dependence on several demographic and geographic factors (age, gender, quarterly season, year, and super-district) in southern Thailand from 1999 to 2004. When the dependent variable is disease count, the

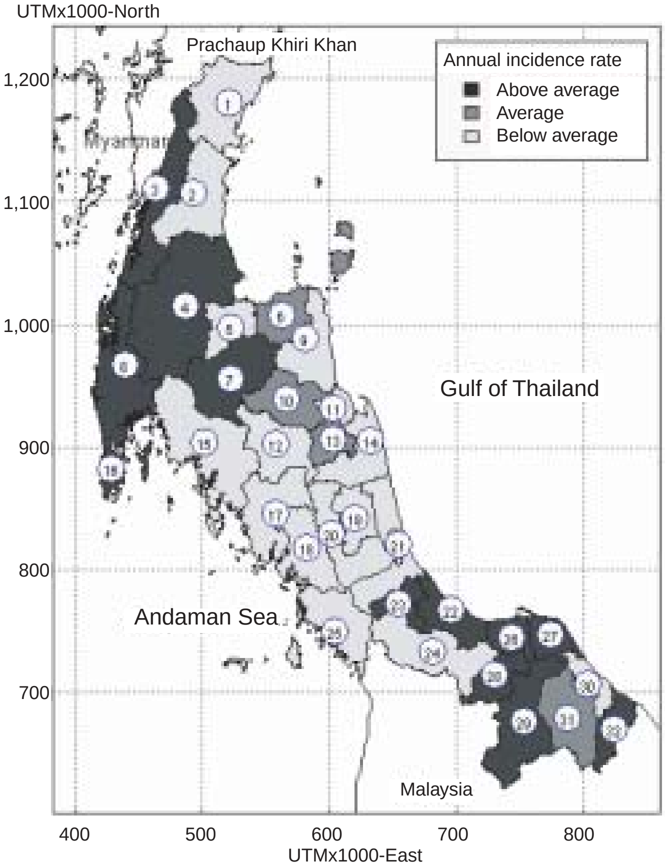


Fig 3–TB incidence patterns in super-districts of southern Thailand.

Poisson and negative binomial generalized linear models are usually considered to be most statistically appropriate. The Poisson distribution assumes events to be independent and does not account for clustering, over-dispersion or serial correlation. A negative binomial GLM is an extension of the Poisson regression model that allows for over-dispersion. Negative binomial models containing gender, age, quarterly season and super-district as factors were fitted to disease incidence. For these data the negative binomial model gave a poor fit, as indicated by the residual plot. Linear regression models containing gender, age, quarterly season and super-district as factors were fitted to the log-transformed disease incidences, with zero cell counts replaced by a constant before log-transformation. This model provided an acceptable fit with an adjusted R-squared of 0.64. The overall annual incidence of TB in the 14 southern provinces of Thailand was 0.45 per 1,000 population.

No gender differences in incidence rates for TB were found in people below age 25 years. However, substantial differences between males and females were observed among those age ≥ 25 years. This is consistent with age and gender patterns found in other recent studies. Similar studies by the World Health Organization (WHO) in Southeast Asia have shown the male:female ratio of incidence rates for TB reported in 2006 were lower in lower age groups, and gradually increased with older age groups, with cases among men being 3.5 times greater than women in older age groups (WHO, 2009b). Epidemiological findings indicate in most settings, TB incidence rates are greater in males of all ages except in childhood, they are higher in females. Studies have reported gender differences in TB incidence begin to appear between 10 and 16 years of age and incidence rates remain higher for males than females thereafter (WHO, 2009b).

During the six-year period of the current study (1999-2004), there were no changes in trend for the incidence of TB. However, the incidence rates were higher during each first quarter (January to March), with lower rates found during the second half of each year. Studies from the UK (Douglas *et al*, 1996) and Spain (Rios *et al*, 2000) also showed seasonal variation in TB rates, with higher levels of TB notifications during the summer months. The cause for the increased number of reported rates during the summer months is unclear. It is hypothesized seasonal fluctuations in vitamin D serum levels may contribute to impaired host defense mechanisms against TB bacilli (Davies, 1997). However, a study in Russia gives contrasting results, with hospital admissions for TB increasing during the winter months (from October to March) and declining during the warmer months (April to September) (Atun *et al*, 2005).

A striking finding in this study was the highest TB incidences occurred in all the super-districts of Pattani and Yala and in one super-district of Narathiwat Province (the 3 southernmost provinces). However, the adjoining super-districts of Ranong, Phang-Nga and Phuket Provinces (on the west coast), north of the 3 southernmost provinces, also showed higher incidences of TB. These findings are consistent with a recent study which found higher incidences on the west coast and in the southernmost provinces (Jittimanee *et al*, 2009). Possible reasons for higher than average incidence rates in these provinces were not investigated in our study, and could be due to HIV, which contributes to high incidence of TB. In Thailand, high TB case-fatality rates have been reported in areas with high HIV rates in the general population (Wibulpolprasert, 2007). Besides this, urban migration and cross-border population movements are also contributing factors to TB epidemics in these areas (World Health Organization, 2005). Our study findings also revealed that high TB incidence occurred in border provinces in Thailand including Ranong, Songkhla, Yala and Narathiwat. Lower socio-economic status, drug addiction and increasing poverty seem to be linked with the re-emergence of TB (Carolyn, 1996) and further research in this area is needed.

Perhaps the major limitation of this study is that the precise burden of TB in Southern provinces Thailand is unknown as that the surveillance data from the Ministry of Public Health is known to be under-reported for major infectious diseases (Lumbiganon *et al*, 1990; Saengwonloey *et al*, 2003; Intusoma *et al*, 2008). It is not based on registration in TB clinics in most public hospitals. Despite this limitation, the findings should reflect the relative patterns of TB incidence with respect to demographic, spatial and temporal variation for the southern provinces of Thailand, even though the absolute extent of these incidence rates is inaccurate.

ACKNOWLEDGEMENTS

We thank the Ministry of Public Health, for providing the data. This study was funded by the Graduate School, Prince of Songkla University. We are grateful to Prof Don McNeil who supervised our research and Dr Petchawan Pungrassami for the helpful comments and suggestions.

REFERENCES

Atun RA, Samyshkin YA, Drobniewski F, *et al*. Seasonal variation and hospital utilization for tuberculosis in Russia: hospitals as social care institutions. *Eur J Pub Health* 2005; 15: 350-4.

Bacaër N, Ouifki R, Pretorius C, *et al*. Modeling the joint epidemics of TB and HIV in a South African township. *J Math Biol* 2008; 57: 557-93.

Carolyn S. Healthy cities or unhealthy island? The health and social implications of urban inequality. *Environ Urban* 1996; 8: 9-30.

Davies PD. Seasonality of tuberculosis. *Thorax* 1997; 52: 398.

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

Dye C, Garnett GP, Sleeman K and Williams BG. Prospects for worldwide Tuberculosis control under the WHO DOTS strategy: Directly Observed Short Course Therapy. *Lancet* 1998; 352: 1886-91.

Intusoma U, Sornsrivichai V, Jiraphongsa C, *et al*. Epidemiology, clinical presentations and burden of rotavirus diarrhea in children under five seen at Ramathibodi Hospital, Thailand. *J Med Assoc Thai* 2008; 91: 1350-5.

Jittimanee S, Vorasingha J, Mad-asin W, *et al*. Tuberculosis in Thailand: epidemiology and program performance, 2001-2005. *Int J Infection Dis* 2009; 13: 436-42.

Legrand J, Sanchez A, Le Pont E, *et al*. Modeling the impact of tuberculosis control strategies in highly endemic overcrowded prisons. *PLoS One* 2008; 3: e2100.

Lumbiganon P, Panamonta M, Laopaiboon M, *et al*. Why are Thai official perinatal and infant mortality rates so low? *Int J Epidemiol* 1990; 19: 997-1000.

Nunes C Tuberculosis incidence in Portugal: spatiotemporal clustering. *Int J Health Geogr* 2007; 6-30.

Nguyen T. Tuberculosis: a global health emergency. *Outlook* 1999; 17: 1-8.

Phomborphub B, Pungrassami P, Boonkitjaroen T. Village health volunteer participation in tuberculosis control in Southern Thailand. *Southeast Asian J Trop Med Public Health* 2008; 39: 542-8.

R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2008. [Cited 2008 Feb 8]. Available from: URL: http://www.R-project.org

Rios M, Garcia JM, Sanchez JA, *et al*. A statistical analysis of the seasonality in pulmonary tuberculosis. *Eur J Epidemiol* 2000; 16: 483-8.

Saengwonloey O, Jiraphongsa C, Foy H. Thailand report: HIV/AIDS surveillance 1998. *J Acquir Immune Defic Syndr* 2003; 32 (suppl 1): S63-7.

Thongraung W, Chongsuvivatwang V, Pungrassamee P. Multilevel factors affecting tuberculosis diagnosis and initial treatment. *J Eval Clin Pract* 2008; 14: 378-84.

Tongkumchum P, McNeil D. Confidence intervals using contrasts for regression model, *Songklanakarin J Sci Technol* 2009; 31: 151-6.

Uplekar MW, Rangan S, Weiss MG, *et al*. Attention to gender issues in tuberculosis control. *Int J Tuberc Lung Dis* 2001; 5: 220-4.

Uthman OA. Spatial and Temporal Variations in Incidence of Tuberculosis in Africa, 1991 to 2005. *World Health Popul* 2008; 10: 5-15.

Venables WN, Ripley BD. Modern applied statistics with S. 4th ed. New York: Springer-Verlag, 2002: 139-208.

Waaler HA, Geser A, Andersen S. The use of mathematical models in the study of the epidemiology of tuberculosis. *Am J Public Health Nations Health* 1962; 52: 1002-13.

Wibulpolprasert S. Thailand health profile 2005-2007. Bangkok: The War Veterans Organization of Thailand, 2007.

Williams BG, Granich R, Chauhan LS, Dharmshaktu NS, Dye C. The impact of HIV AIDS on the control of tuberculosis in India. *PNAS* 2005; 102: 9619-24.

World Health Organization. Overview of Thai/Myanmar boarder health situation 2005. Bangkok; WHO country office for Thailand, 2005.

World Health Organization (WHO). Tuberculosis control in the South East Asia Region: WHO annual report 2009. Geneva: WHO, 2009a.

World Health Organization (WHO). Global tuberculosis control 2009: epidemiology, strategy, financing: WHO report 2009. Geneva: WHO, 2009b.

World Health Organization (WHO). Global tuberculosis control: surveillance, planning, financing: WHO report 2008. Geneva: WHO, 2008.