Spatiotemporal characteristics and the epidemiology of tuberculosis in China from 2004 to 2017 by the nationwide surveillance system

Zhongbao Zuo1*, Miaochan Wang1*, Huaizhong Cui1*, Ying Wang1, Jing Wu1, Jianjiang Qi1, Kenv Pan1, Dongming Sui1, Pengtao Liu2, Aifang Xu1#

1 Department of clinical laboratory, Hangzhou Xixi Hospital, 310023, Zhejiang Province, China.

2 Department of General Courses, Weifang Medical University, 261053, Shandong Province, China.

*These authors contributed equally to this work.

#Corresponding author: Aifang Xu, 2 Hengbu Road, Xihu District, Zhejiang 310023, China. Phone: +86-10-86481591; Fax: +86-10-86481822; Email: 13616500869@163.com.

Abstract

Background

China has always been one of the countries with the most serious tuberculosis epidemic in the world. Our study was to observe the Spatial-temporal characteristics and the epidemiology of tuberculosis in China from 2004 to 2017 with Joinpoint regression analysis, Seasonal Autoregressive integrated moving average (SARIMA) model, geographic cluster, and multivariate time series model.

Methods

The data of TB from January 2004 to December 2017 were obtained from the notifiable infectious disease reporting system supplied by the Chinese Center for Disease Control and Prevention. The incidence trend of TB was observed by the Joinpoint regression analysis. The Seasonal autoregressive integrated moving average (SARIMA) model was used to predict the monthly incidence. Geographic clusters was employed to analyze the spatial autocorrelation. The relative importance component of TB was detected by the multivariate time series model.

Results

We included 13,991,850 TB cases from January 2004 to December 2017, with a yearly average morbidity of 999,417 cases. The final selected model was the 0 Joinpoint model (P=0.0001) with

an annual average percent change (AAPC) of -3.3 (95% CI: -4.3 to -2.2, P<0.001). A seasonality was observed across the fourteen years, and the seasonal peaks were in January and March every year. The best SARIMA model was (0, 1, 1) X $(0, 1, 1)_{12}$ which can be written as (1-B) $(1-B^{12})$ X_t = (1-0.42349B) $(1-0.43338B^{12})$ ε_t, with a minimum AIC (880.5) and SBC (886.4). The predicted value and the original incidence data of 2017 were well matched. The MSE, RMSE, MAE, and MAPE of the modelling performance were 201.76, 14.2, 8.4 and 0.06, respectively. The provinces with a high incidence were located in the northwest (Xinjiang, Tibet) and south (Guangxi, Guizhou, Hainan) of China. The hotspot of TB transmission was mainly located at southern region of China from 2004 to 2008, including Hainan, Guangxi, Guizhou, and Chongqing, which disappeared in the later years. The autoregressive component had a leading role in the incidence of TB which accounted for 81.5% - 84.5% of the patients on average. The endemic component was about twice as large in the western provinces as the average while the spatial-temporal component was less important there. Most of the high incidences (>70 cases per 100,000) were influenced by the autoregressive component for the past fourteen years.

Conclusion

In a word, China still has a high TB incidence. However, the incidence rate of TB was significantly decreasing from 2004 to 2017 in China. Seasonal peaks were in January and March every year. Obvious geographical clusters were observed in Tibet and Xinjiang Province. The relative importance component of TB driving transmission was distinguished from the multivariate time series model. For every provinces over the past fourteen years, the autoregressive component played a leading role in the incidence of TB which need us to enhance the early protective implementation.

Keywords: Tuberculosis; spatial-temporal; epidemiology; multivariate time series model

Background

Tuberculosis (TB) continues to challenge the international community. It is estimated that there are about 1.7 billion people with potential TB infection, accounting for 23% of the world's population, are at risk of developing TB disease during their lifetime[1]. Moreover, the global burden

was estimated by the World Health Organization (WHO) at 10.0 million incident cases in 2017. It is also one of the top 10 causes of death which caused an estimated 1.6 million deaths in 2017, and has killed more people than other infectious diseases in the past few decades [1, 2].

China has always been one of the countries with the most serious tuberculosis epidemic in the world [3-6]. There were 866,000 patients with infection of TB in China, 2018[7]. Due to the continuous attention to public health and increasing investment in resources, China's tuberculosis epidemic has significantly improved in recent years. However, due to the large number of people infected with TB, the epidemic situation of tuberculosis is still not optimistic, so further long-term research on the incidence of it in China is needed.

Currently, China has conducted five national epidemiological investigations to find the epidemiological characteristics of Tuberculosis. However, the spatiotemporal distributions of Tuberculosis cannot be evaluated continuously, and the survey was unable to measure other important indicators of the severity of the epidemic. The mathematical models may help us better understand the epidemiological characteristics of Tuberculosis. Some of the studies mainly focused on the seasonality impact on the transmission of Tuberculosis [5, 8, 9], while others focused on the spatial distributions [10, 11]. There is no model that assesses the spatiotemporal characteristics and the epidemiology of tuberculosis among the whole population in China over fourteen years.

The aim of this study was to observe the Spatial-temporal characteristics and the epidemiology of tuberculosis in China from 2004 to 2017. The incidence trend of the TB was observed by the Joinpoint regression analysis. The Seasonal autoregressive integrated moving average (SARIMA) model was used to predict the monthly incidence. Geographic clusters was employed to analyze the spatial autocorrelation. The relative importance component of TB was detected by the multivariate time series model. These models additively divided TB risks into spatiotemporal, autoregressive, and endemic components.

Methods

The data collection

Tuberculosis incidence data were extracted from the Chinese Center for Disease Control and Prevention (http://www.phsciencedata.cn/Share/edtShareNew.jsp?id=39208) in 31 provinces of China from 2004 to 2017. The data were aggregated to 168 monthly counts across the fourteen years. Population data came from the website of the statistical yearbook of the National Bureau of Statistics (http://www.stats.gov.cn/tjsj/ndsj/). The population size was easy to find in the website, and it represented the average population each year.

Joinpoint regression

From 2004 to 2017, the continuous change of the TB incidence trend was analyzed using Joinpoint software. The grid search method was applied to find significant trends, and multiple permutation tests were applied to detect the Joinpoint points for each trend [12-14]. The overall time trend was calculated by the annual average rate of change (AAPC). If the final model was 0 Joinpoint model, the average percent change (APC) was considered equal to AAPC. We used the Joinpoint regression model to find the long-term trend of the TB incidence.

Time-series estimation

The SARIMA model was used to predict the future trends in many disease incidences [13, 15-17]. In our study, A SARIMA model was applied to predict the incidence of TB epidemics in China. The SARIMA model can be written as the form of (p, d, q) (P, D, Q) [s], which P, D, and Q indicate seasonal SAR terms, seasonal differencing, and seasonal SMA terms, respectively; p, d, and q indicate non-seasonal AR terms, non-seasonal differencing, and non-seasonal MA, respectively; s indicated the seasonal period (s = 12 in our study).

The construction of the SARIMA model can be divided into the following steps. First, an augmented Dickey-Fuller (ADF) test was performed to test the stationary status of time series. Second, model parameters (p, d, q, P, D, and Q) were determined by autocorrelation function (ACF) plot, partial autocorrelation function (PACF) plot, and inverse autocorrelation function (IACF) plot. An alternative SARIMA model was constructed by transforming the parameters of model. Lastly, the Akaike information criterion (AIC) and Schwartz Bayesian Criterion (SBC) were used to determine the fitness of different SARIMA models. An optimal model was considered to have the lowest AIC and SBC values, and the residuals of the final model were tested by the Box-Ljung test to know whether they were time independent. The mean square error (MSE), mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE) were used to see the predictive validity of the models. We use year 2004-2016 to construct the SARIMA model, and year 2017 to testify the forecast of the model. The SARIMA model is used to forecast the short-term incidence of TB to testify the accuracy of model. We also decompose the monthly data into the overall trend, seasonal trend, and random noise with a goal to identify the truly long-term trend.

Spatial autocorrelation analysis

Spatial analysis was used to identify the clustering regions and observe geographic variation [18, 19]. Global Moran's I of reported TB cases was computed to detect the spatial clustering pattern. A Moran's I value is between -1 and 1, whereas the value near 1 means positive spatial autocorrelation, the value near −1 means negative spatial autocorrelation, and 0 means random distribution. Local Moran's Index was calculated and a hotspot analysis was performed to determine the location of clusters. Local Moran's Index was applied to determine the spatial autocorrelation, which detects some spatial clusters with similar adjacent features and exception values. When the incidences rate had similar low values or high values, these areas were deemed as having positive autocorrelation (low-low or high-high autocorrelation). If not, they were defined as having a negative autocorrelation (low-high or high-low autocorrelation) [10].

The multivariate time series model

A multivariate time-series model for disease counts $Y_{i,t}$ during periods t = 1, …, T from units i = 1, … , I was first established by Held et al[20] and was extended and applied in some papers [21-23]. The $Y_{i,t}$ denoted the number of TB cases which were considered to be a negative binomial distribution $Yit|Y_{i,t-1} \sim NegBin(u_{it}, \psi)$, with an additively decomposed mean:

$$u_{it} = \upsilon_{it} e_{it} + \lambda_{it} Y_{i,t-1} + \phi_{it} \sum_{j \neq i} w_{ji} Y_{j,t-1},$$

Where ψ is an over-dispersed parameter that the conditional variance of $Y_{i,t}$ is $\mu_{it}(1 + \psi u_{it})$. $\upsilon_{it} e_{it}$ is the endemic component, and the autoregressive component $\lambda_{it} Y_{i,t-1}$ reflects the patient numbers at previous time. The spatiotemporal component $\phi_{it} \sum_{j \neq i} w_{ji} Y_{j,t-1}$ reflects the transmission among different units. Each parameter υ_{it}, λ_{it}, and ϕ_{it} follow the form of log-linear:

$$\log(\upsilon_{it}) = \alpha^{(v)} + b_i^{(v)} + \sum_{s=1}^{S} \{\gamma \sin(w_s t) + \delta \cos(w_s t)\},$$

$$\log(\lambda_i) = \alpha^{(\lambda)} + b_i^{(\lambda)},$$

$$\log(\phi_i) = \alpha^{(\phi)} + b_i^{(\phi)},$$

Where $\alpha^{(v)}$, $\alpha^{(\lambda)}$ and $\alpha^{(\phi)}$ are intercepts and $b_i^{(v)}$, $b_i^{(\lambda)}$ and $b_i^{(\phi)}$ are random effects accounting for heterogeneity among different regions. The endemic υ_{it} contains a sinusoidal frequency wave ($w_s = 2\pi/12$ for monthly data), and S is the seasonal parameters. The population fraction e_i can be used as a multiplicative offset for the regional specific measure for the incidence of infectious disease.

The weights ω_{ji} describe the transmission from district j to district I. Considering that most regions are very large, higher-order neighbourhood are not that relevant as we only constructed our model with first-order neighbourhood. The score rule of the Dawid-Sebastiani score ("dss") was applied to identify the optimal model with random effects. The optimal model corresponds to lower scores with better predictions [22, 24]. All the multivariate time analysis used the R package Surveillance.

Statistical analysis

The incidence trend of TB from 2004 to 2017 was observed by the Joinpoint software (version 4.7.0.0). The Seasonal autoregressive integrated moving average (SARIMA) model was used to predict the monthly incidence of TB by SAS9.4 (SAS Institute Inc., Cary, NC). Geographic clusters was employed to analyze the spatial autocorrelation with ArcGIS software (version 10.2, ESRI Inc.; Redlands, CA, USA). The relative importance component of TB was detected by the multivariate time series model with R software (version 3.6.0, package = surveillance). P value < 0.05 was considered as statistically significant for all the tests.

Results

Time trends, seasonal characteristics of the TB incidence

We included 13,991,850 TB cases from January 2004 to December 2017, with a yearly average morbidity of 999,417 cases. A fluctuant reduction was seen from 74.57 (/100,000) cases in 2004 to 60.08 (100,000) cases in 2017, with the highest incidence of 96.30 (100,000) cases in 2005. The final model was the 0 Joinpoint model (P=0.18). The annual average percent change (AAPC) was -3.3 (95% CI: -4.3 to -2.2, P<0.001) from 2004 to 2017, indicating a downward trend in the TB incidence (Fig 1).

The occurrence of TB with obvious seasonality was observed in the past fourteen years (Fig 2), and the seasonal cycle kept on fluctuating within 12 months. There were two incidence peaks in January and March every year, with a burst From December of the previous year to January of the following year.

The null hypothesis of white noise was strongly rejected with the results of the white noise test (χ^2 = 131.98, DF = 6, P<0.0001), which can extract some useful information from the time series. Although the null hypothesis was significant (Tau = -3.91, P=0.003, lag = 1) for the augmented

Dickey-Fuller (ADF) test, we should make a seasonal difference taking account of the fluctuation of the incidence figure. We performed a seasonal differencing to make sure that the transformed TB incidence was stationary (Tau = -7.6, P<0.0001, lag = 1) to better construct the SARIMA model (Fig 3). Based on the figures of PACF, ACF, and IACF, the best ARIMA model was (0, 1, 1) X (0, 1, 1)$_{12}$ which can be written as (1-B) (1-B^{12}) X_t = (1-0.42349B) (1-0.43338B^{12}) ε_t, with a minimum AIC (880.5) and SBC (886.4). There was no significant correlation between residuals (lag= 6, χ^2= 3.65, DF = 3, P=0.45), and the residual was a white noise. We then did an incidence forecast of 2017 shown in Fig 2, the predicted and actual incidence were shown in table 1. The predicted value and the original incidence data of 2017 were well matched. The mean square error (MSE), mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute error (MAE) of the modelling performance were 201.76, 0.06, 14.2, and 8.4 respectively. The time series can divide into three components: seasonal effect, trend curve, and irregular noise. The seasonal effect refers to the fluctuations of the trend that is reproduced in a similar way every year, the trend curve is the long-term movement of the time series, and the irregular noise is the surplus component after trend curve and seasonal effect are removed. After eliminating the influence of seasonal effect and irregular noise on TB, the incidence curve of TB became smoother (Fig 2), and it was found that the trend of the incidence from 2004 to 2016 was gradually decreasing.

Spatial clustering distribution and geographic characteristics

The TB cases were reported in every province of China from 2004 to 2017, with the lowest incidence of 19.52(/100,000) in Hebei Province (2015) to the highest incidence of 204.45(/100,000) in Xinjiang Province (2005). Xinjiang Province was the most prevalent province of tuberculosis in China from 2004 to 2017, and the incidence of Tibet was in a high level since 2012 (Fig 4). The provinces with a high incidence were located in the northwest (Xinjiang, Tibet) and south (Guangxi, Guizhou, Hainan) of China.

Based on the global autocorrelation analysis, the distribution of TB was spatially correlated from 2004 to 2017 (Table 2). The Moran's index range from 0.28 to 0.36, and had the highest index in 2011(Moran's index = 0.36, Z-score = 5.51, P<0.001). According to the local Moran's I autocorrelation results, it was found that there were totally 35 high-high clusters and 1 high-low cluster from 2004 to 2017 (Table 3), with 4, 3, 3, 3, 5, 2, 2, 2, 2, 2, 2, 2, 2, and 2 clusters each year. The hotspot of TB transmission was mainly located at southern region of China from 2004 to 2008,

including Hainan, Guangxi, Guizhou, and Chongqing, which disappeared in the later years. It should be noted that the center of the high-high clusters moved from the East to the Northwest (Xinjiang and Tibet) after 2008, and Tibet was a high-low cluster in 2008.

Multivariate time series analysis

Two models following negative binomial distribution and the Poisson distribution constructed by the monthly data from 2004 to 2017were built in the first step, and the AIC of the two models were 72247.32 and 260511.37, which meant the better distribution of the model would be the negative binomial distribution. Second, we included the random effects of the model, and found that the random effects model (0.20) introduced by DSS rule was better than the negative binomial distribution model (2.06). Considering that most regions are very large, higher-order neighbourhood are not that relevant when we only construct the model with first-order neighbourhood.

In order to classify the spatial-temporal effect of the TB, the relative importance of the model components by province, with an average of fourteen years is shown in Fig 6. The autoregressive component had a leading role in the incidence of TB which accounted for 81.5% - 84.5% of the patients across all provinces on average (Fig 6B). The endemic component was about twice as large in the western provinces as the average while the spatial-temporal component was less important there (Fig 6A/C). It should be noted that some economic circles, such as the Yangtze River Delta economic circle (Zhejiang, Jiangsu, and shanghai), Pearl River Delta economic circle (Guangxi and Guangdong), Bohai Economic Rim (Hebei, Tianjin, Beijing, and Shanxi) and Hanjiang ecological economic belt (Henan and Hubei), had higher proportions of the spatial-temporal component (especially in Beijing), whereas there was very little spatial correlation in the western provinces.

An intuitive method to quantify the relative contributions of the high incidence regions (>70 cases per 100,000 persons over fourteen years) of the three components is provided by Fig 7. In general, most of the high incidences were mainly affected by the autoregressive component for the past fourteen years. There was clear seasonality with two incidence peaks in January and March every year, with a burst From December of the previous year to January of the following year. Guangxi, Heilongjiang, Hubei, Guangdong and Hainan were partly affected by the spatial-temporal component, while the rest of the high incidence provinces had nearly no associations with the spatial-temporal effect.

Discussion

According to our research, there were 13,991,850 TB cases from January 2004 to December 2017, with a yearly average morbidity of 999,417 cases which was a huge burden for the public health of China. Understanding the epidemiology patterns of TB may help China to reduce the number of TB cases which ranked second in 2017 according to the WHO report [1]. The incidence of TB from 74.58 (/100,000) cases in 2004 to 60.08 (/100,000) cases in 2017 which was a 19.4% reduction of TB incidence. The annual average percent change (AAPC) was -3.3, which is better than the world average of 2% [1]. The reason for the decline of TB incidence is the rising GDP (Gross Domestic Product) (China ranked second in 2019), high urbanization, and the widespread modern control strategy. Previous studies demonstrated that the TB incidence of China decreased with the rising of GDP and better healthy treatment and management [25, 26], which was also found in other countries [27, 28].

Consistent with previous research [29], we found two peaks in January and March every year for TB incidence in China, with close numbers in these two peaks. The low number of confirmed cases in February may probably attribute to the Chinese traditional Spring Festival holiday. The average time from disease onset to confirmation of the diagnosis was 72 days when some infected persons develop active tuberculosis, and patients were most likely to be diagnosed 2-3 months after symptom onset [8]. So, we should enhance patient control and the prevention of susceptible population in the autumn and winter, and the detection of TB in spring.

For a long time, the hotspots were distributed in the northwest areas such as Xinjiang Province and Tibet. Xinjiang Province has been at a high incidence level in the fourteen years, while the incidence in Tibet increased since 2012. Except for 2004, 2010, and 2014, Guizhou Province has been at a high incidence level in the later eleven years. Some provinces such as Hainan, Guangxi, and Chongqing were at a high incidence level before 2009, but have been at a low level since 2009. More attention is needed in these high incidence areas, especially in Xinjiang, Tibet and Guizhou, which may need more financial assistance. It should be noted that some High-High spatial autocorrelation including Hainan, Guizhou, Guangxi, and Chongqing Province have disappeared since 2009, while Xinjiang Province and Tibet have become new H-H regional areas since 2009. The possible explanation is the unbalanced economic development in these areas [30]. Some studies [31-33] have demonstrated that there has positive correlations between the poverty level of regions, families or individuals and the incidence of tuberculosis.

At the average level of the province component over the fourteen years, autoregressive components dominated all the provinces which can explain 81.5%-84.5% of the incidence, while the spatiotemporal component was mainly located in the well-developed provinces. For some provinces such as Beijing, Jiangsu and other well-developed economic provinces which were partly affected by the spatiotemporal component, it is recommended to monitor TB infection of the floating population form the neighbouring areas. For example, individuals who work in Beijing but become infected with TB in their hometowns should stay at home before anti-tuberculosis treatment and maintain the treatment for a couple of weeks, avoiding going to public places or having close contact with others. We also did an analysis for the provinces with a high incidence (>70 cases per 100,000 over fourteen years) of the three components. For the autoregressive component which dominated all the high incidence provinces, early protective implementation 2-3 months ahead of the peak could help us reduce the number of TB patients [8]. For the endemic parts, most infected patients could be explained by living conditions, ecological and climatological changes, and socioeconomic activities. Active treatment for TB patients and cutting off the pathway of transmission may be the most effective way to prevent TB [8, 34]. Another important method is increasing the public awareness, especially among old people and children, and enhancing their physical exercise, immunity, and general hygiene. In addition, the spatial-temporal component can also affect the transmission of TB. Guangxi and Guangdong Provinces, which are in the south-east coastal area, were partly influenced by the spatial-temporal component, indicating that these regions may have imported TB from adjacent country with high incidence such as Philippines [35, 36] or the neighbouring province Guizhou. Alarmingly, although there was no clear evidence that Tibet and Xinjiang had a high value of spatial-temporal component, we still need to pay attention to transmission from India [34, 35] which was ranked first in global TB patients.

Our study had several limitation. First, the monthly data from 2004 to 2017 did not collect some risk factors including socioeconomic status, climatic factors, gender, age, and human activities. The relationship between the incidence of TB and these factors was still unknown. These factors should be included in the future studies in order to get an accurate multivariate time series model. Second, we included TB patients reported from the passive surveillance system which inevitably underestimated the total number of TB cases. Further researches could consider the level of reporting, including some subclinical and mild individuals not accessing healthcare. Lastly, the level

of diagnosis in some provinces can lead to an underestimation of the TB incidence. We should think over the diagnostic level in the future studies to correct the incidence.

Conclusion

In conclusion, China still has a high TB incidence. However, the incidence rate of TB was significantly decreasing from 2004 to 2017 in China. Seasonal peaks were in January and March every year, with a burst From December of the previous year to January. Obvious geographical clusters were observed in Tibet and Xinjiang Province. The relative importance component of TB driving transmission was distinguished from the multivariate time series model. For every provinces over the past fourteen years, the autoregressive component played a leading role in the incidence of TB which need us to enhance the early protective implementation.

Abbreviation

Tuberculosis (TB)

Autoregressive integrated moving average (SARIMA)

Annual average percent change (AAPC)

Autocorrelation function (ACF)

Partial autocorrelation function (PACF)

Inverse autocorrelation function (IACF)

Akaike information criterion (AIC)

Schwartz Bayesian Criterion (SBC)

Mean square error (MSE),

Mean absolute percentage error (MAPE),

Mean absolute error (MAE),

Root mean square error (RMSE)

Declarations

Ethics approval and consent to participate

The ethical approval is not warranted for our present work as the monthly monitoring data of TB morbidity are publicly available in China.

Consent for publication

Yes

Availability of data and materials

All data were enclosed to the online supplementary materials.

Competing interests

None declared.

Funding

This study was supported by grants from the Hangzhou Science and Technology Bureau (grant#91203B11), the Natural Science Foundation of Zhejiang Province (grant #LGF19H19003), and Zhejiang medical and health science and technology plan (NO. 2019KY533). The fund providers had no roles in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

Conceived and designed the research: JW, DS, AX. Data collection: HC, KP, JQ. Data analysis: ZZ, HC. Wrote the paper: ZZ, YW. Reviewed and revised the paper: ZZ, JW. All authors read and approved the final manuscript.

Acknowledgements

We thank all the clinical personnel for investigating and reporting the information about the TB cases. Thanks to Dr. Sebastian Meyer, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Institute for Medical informatics, Biometry and Epidemiology (IMBE), for help with the R package surveillance. Thanks to Dr. Edward C. Mignot, Shandong University, for linguistic advice.

Reference

[1] WHO. Global tuberculosis report 2018 [M]. 2018.

[2] MOOSAZADEH M, KHANJANI N, NASEHI M, et al. Predicting the Incidence of Smear Positive Tuberculosis Cases in Iran Using Time Series Analysis [J]. Iranian journal of public health, 2015, 44(11): 1526-34.

[3] ZHENGHONG R. The Temporal Characteristics and Trend of Tuberculosis Incidence Cases in China Based on a National Surveillance Data since 2005 [J]. Chinese Journal of Health Statistics, 2013, 02(30): 158-61.

[4] XIE C, XU L, WANG X, et al. [Epidemiological characteristics and spatial-temporal clustering analysis on pulmonary tuberculosis in Changsha from 2013 to 2016] [J]. Zhong nan da xue xue bao Yi xue ban = Journal of Central South University Medical sciences, 2018, 43(8): 898-903.

[5] WANG Y, XU C, ZHANG S, et al. Temporal trends analysis of tuberculosis morbidity in mainland

China from 1997 to 2025 using a new SARIMA-NARNNX hybrid model [J]. BMJ open, 2019, 9(7): e024409.
[6] WANG H, TIAN C W, WANG W M, et al. Time-series analysis of tuberculosis from 2005 to 2017 in China [J]. Epidemiology and infection, 2018, 146(8): 935-9.
[7] WHO. China-Tuberculosis profile [M]. 2019.
[8] GUO Z, XIAO D, WANG X, et al. Epidemiological characteristics of pulmonary tuberculosis in mainland China from 2004 to 2015: a model-based analysis [J]. BMC public health, 2019, 19(1): 219.
[9] YANG Y, GUO C. Seasonality Impact on the Transmission Dynamics of Tuberculosis [J]. Computational and mathematical methods in medicine, 2016, 2016(8713924.
[10] YANG S, GAO Y, LUO W, et al. Spatiotemporal Distribution of Tuberculosis during Urbanization in the New Urban Area of Nanchang City, China, 2010-2018 [J]. International journal of environmental research and public health, 2019, 16(22):
[11] CHEN J, QIU Y, YANG R, et al. The characteristics of spatial-temporal distribution and cluster of tuberculosis in Yunnan Province, China, 2005-2018 [J]. BMC public health, 2019, 19(1): 1715.
[12] KIM H J, FAY M P, FEUER E J, et al. Permutation tests for joinpoint regression with applications to cancer rates [J]. Statistics in medicine, 2000, 19(3): 335-51.
[13] WU H, WANG X, XUE M, et al. Spatial-temporal characteristics and the epidemiology of haemorrhagic fever with renal syndrome from 2007 to 2016 in Zhejiang Province, China [J]. Scientific reports, 2018, 8(1): 10244.
[14] YANG S, WU J, DING C, et al. Epidemiological features of and changes in incidence of infectious diseases in China in the first decade after the SARS outbreak: an observational trend study [J]. The Lancet Infectious diseases, 2017, 17(7): 716-25.
[15] LI S, CAO W, REN H, et al. Time Series Analysis of Hemorrhagic Fever with Renal Syndrome: A Case Study in Jiaonan County, China [J]. PloS one, 2016, 11(10): e0163771.
[16] SONG Y, WANG F, WANG B, et al. Time series analyses of hand, foot and mouth disease integrating weather variables [J]. PloS one, 2015, 10(3): e0117296.
[17] GAUDART J, TOURE O, DESSAY N, et al. Modelling malaria incidence with environmental dependency in a locality of Sudanese savannah area, Mali [J]. Malaria journal, 2009, 8(61.
[18] LIU K, CAI J, WANG S, et al. Identification of Distribution Characteristics and Epidemic Trends of Hepatitis E in Zhejiang Province, China from 2007 to 2012 [J]. Scientific reports, 2016, 6(25407.
[19] LIU Z, SHI O, YAN Q, et al. Changing epidemiological patterns of HIV and AIDS in China in the post-SARS era identified by the nationwide surveillance system [J]. 2018, 18(1): 700.
[20] HELD L, H HLE M, HOFMANN M. A statistical framework for the analysis of multivariate infectious disease surveillance data [J]. Statist Modllng, 2005, 5(
[21] PAUL M, HELD L. Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts [J]. Statistics in medicine, 2011, 30(10): 1118-36.
[22] MEYER S, HELD L, H HLE M. Spatio-Temporal Analysis of Epidemic Phenomena Using the R Package surveillance [M]. 2014.
[23] HELD L, PAUL M. Modeling seasonality in space-time infectious disease surveillance data [J]. Biometrical journal Biometrische Zeitschrift, 2012, 54(824-43.
[24] MEYER S, HELD L. Power-law models for infectious disease spread [J]. The Annals of Applied Statistics, 2014, 8(1612-39.
[25] MAO W, JIANG W, HAMILTON C, et al. Over- and under-treatment of TB patients in Eastern China: an analysis based on health insurance claims data [J]. Tropical medicine & international health : TM & IH, 2019, 24(9): 1078-87.

[26] DU J, EMILIO D, PANG Y, et al. Tuberculosis Hospitalization Fees and Bed Utilization in China from 1999 to 2009: The Results of a National Survey of Tuberculosis Specialized Hospitals [J]. PloS one, 2015, 10(10): e0139901.
[27] FUADY A, HOUWELING T A J, MANSYUR M, et al. Effect of financial support on reducing the incidence of catastrophic costs among tuberculosis-affected households in Indonesia: eight simulated scenarios [J]. Infectious diseases of poverty, 2019, 8(1): 10.
[28] CHENG J, ZHANG H, ZHAO Y L, et al. Mutual Impact of Diabetes Mellitus and Tuberculosis in China [J]. Biomedical and environmental sciences : BES, 2017, 30(5): 384-9.
[29] YANG Y, GUO C, LIU L, et al. Seasonality Impact on the Transmission Dynamics of Tuberculosis [J]. 2016, 2016(8713924.
[30] LI X X, WANG L X, ZHANG H, et al. Spatial variations of pulmonary tuberculosis prevalence co-impacted by socio-economic and geographic factors in People's Republic of China, 2010 [J]. BMC public health, 2014, 14(257.
[31] MARMOT M, ALLEN J, BELL R, et al. WHO European review of social determinants of health and the health divide [J]. Lancet (London, England), 2012, 380(9846): 1011-29.
[32] NEWMAN L, BAUM F, JAVANPARAST S, et al. Addressing social determinants of health inequities through settings: a rapid review [J]. Health promotion international, 2015, 30 Suppl 2(ii126-43.
[33] VASSALL A, SIAPKA M, FOSTER N, et al. Cost-effectiveness of Xpert MTB/RIF for tuberculosis diagnosis in South Africa: a real-world cost analysis and economic evaluation [J]. The Lancet Global health, 2017, 5(7): e710-e9.
[34] RAO V G, MUNIYANDI M, BHAT J, et al. Research on tuberculosis in tribal areas in India: A systematic review [J]. The Indian journal of tuberculosis, 2018, 65(1): 8-14.
[35] RAGONNET R, TRAUER J M, GEARD N, et al. Profiling Mycobacterium tuberculosis transmission and the resulting disease burden in the five highest tuberculosis burden countries [J]. BMC medicine, 2019, 17(1): 208.
[36] KIM S, DE LOS REYES A A T, JUNG E. Mathematical model and intervention strategies for mitigating tuberculosis in the Philippines [J]. Journal of theoretical biology, 2018, 443(100-12.

Figure Legends

Fig 1. Trend of TB incidence rate from 2004 to 2017 shown by the Joinpoint software. The red squares denote the incidence of each year and the blue line is the slope of the annual percent change (APC).

Fig 2. The actual and seasonal-adjusted incidence of TB in China, from January 2004 to December 2017 at monthly intervals. The blue line is the original incidence, the red line is the seasonal-adjusted incidence, and the green line is the trend line.

Fig 3. The time series of one step of 12 months difference and its three kinds of autocorrelation function plot. (A) The time series after one-step seasonal differences. The x-axis is the time and the y-axis is the difference between the value of incidence and the value a lag of 12 months. The plot (B–D) shows the degree of correlations with past values of the time series. For the plot (B–D), the x-axis is the number of periods of the lag, the y-axis is the coefficient of the autocorrelation, partial autocorrelation, and inverse autocorrelation, respectively. The blue shadows are the boundaries of confidence intervals (two times the standard deviation) of the coefficient. (B) The figure of the autocorrelation of the time series. (C) The

figure of the partial autocorrelation of the time series. (D) The figure of the inverse autocorrelation of the time series.

Fig 4. Maps of the incidence of TB in China, 2004–2017. Maps were created by ArcGIS software (version 10.1, ESRI Inc.; Redlands, CA, USA).

Fig 5. Maps of the local autocorrelation analysis of the incidence rate of TB in China, 2004–2017 by the local Moran's I. Maps were created by ArcGIS software (version 10.1, ESRI Inc.; Redlands, CA, USA). The HH is the high-high spatial autocorrelation, the HL is the high-low spatial autocorrelation, the LH is the low-high spatial autocorrelation, and the LL is the low-low spatial autocorrelation.

Fig 6. The three components of TB on average of fourteen years in the multivariate time series model. This map was created by R software (version 3.3.1, http://www.r-project.org/). The colors represented the value of the proportion of the three components at the province level.

Fig 7. Fitted components in the multivariate time series model for the 12 counties with more than 70 cases during the past fourteen years. The black dots represent the monthly counts of incidence, the light grey area shows the endemic component, the blue area shows the autoregressive component, and the yellow area corresponds to the spatiotemporal component.