



OPEN ACCESS

EDITED BY

Chiara de Waure,
University of Perugia, Italy

REVIEWED BY

Barathan Muttiah,
University of Malaya, Malaysia
Santosh Chokkakula,
Chungbuk National University, Republic of
Korea

*CORRESPONDENCE

Bin Luo
✉ luob@lzu.edu.cn

[†]These authors have contributed equally to
this work

RECEIVED 22 May 2024

ACCEPTED 19 December 2024

PUBLISHED 07 January 2025

CITATION

Guo J, Liu C, Liu F, Zhou E, Ma R, Zhang L and
Luo B (2025) Tuberculosis disease burden in
China: a spatio-temporal clustering and
prediction study.
Front. Public Health 12:1436515.
doi: 10.3389/fpubh.2024.1436515

COPYRIGHT

© 2025 Guo, Liu, Liu, Zhou, Ma, Zhang and
Luo. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Tuberculosis disease burden in China: a spatio-temporal clustering and prediction study

Jingzhe Guo^{1†}, Ce Liu^{1†}, Fang Liu^{2†}, Er kai Zhou¹, Runxue Ma¹,
Ling Zhang¹ and Bin Luo^{1*}

¹Institute of Occupational Health and Environmental Health, School of Public Health, Lanzhou
University, Lanzhou, China, ²Gansu Provincial Center for Disease Prevention and Control, Lanzhou,
China

Introduction: The primary aim of this study is to investigate and predict the prevalence and determinants of tuberculosis disease burden in China. Leveraging high-quality data sources and employing a methodologically rigorous approach, the study endeavors to enhance our understanding of tuberculosis control efforts across different regions of China. First, through nationwide spatio-temporal cluster analysis, we summarized the status of tuberculosis burden in various regions of China and explore the differences, thereby providing a basis for formulating more targeted tuberculosis prevention and control policies in different regions; Subsequently, using a time series-based forecasting model, we conducted the first-ever national tuberculosis burden trend forecast to offer scientific guidance for timely adjustments in planning and resource allocation. This research seeks to contribute significantly to China's existing tuberculosis prevention and control system.

Materials and methods: This research draws upon publicly available pulmonary tuberculosis (PTB) incidence and mortality statistics from 31 provinces and municipalities of mainland China between 2004 and 2018. We organized and classified these data according to province, month, year, and patient age group. Overall, the sample included 14,816,329 new instances of PTB and 42,465 PTB-related fatalities. We used spatiotemporal cluster analysis to record the epidemiological characteristics and incidence patterns of PTB during this period. Additionally, a time series model was constructed to forecast and analyze the incidence and mortality trends of PTB in China.

Results: This study reveals significant regional variations in PTB incidence and mortality in China. Tibet (124.24%) and Xinjiang (114.72%) in western China exhibited the largest percentage change in tuberculosis (TB) incidence, while Zhejiang Province (−50.45%) and Jiangsu Province (−51.33%) in eastern China showed the largest decreases. Regions with significant percentage increases in PTB mortality rates (>100%) included four western regions, six central regions, and five eastern regions. The regions with relatively large percentage decreases in the mortality rate of PTB include Tianjin (−52.25%) and Shanghai (−68.30%). These differences are attributed to two main factors: (1) economic imbalances leading to poor TB control in underdeveloped areas, and (2) differences in TB-related policies among provinces causing uneven distribution of disease risks. Consequently, China may still face challenges in achieving the World Health Organization's 2030 tuberculosis control goals. Nationwide, the mortality rate of PTB in China increased between 2004 and 2018 (percentage change: 105.35%, AAPC: 4.1), while the incidence of PTB showed a downward trend (percentage change: −20.59%, AAPC: −2.1). Among different age groups, the

0–19 age group has the smallest disease burden. While incidence and mortality from TB were primarily found in adults 60 years of age or older, the age group of 0–19 years has the smallest burden of TB, highlighting obvious differences in age characteristics. It is predicted that the mortality rate of TB in China will continue to increase. In summary, the TB epidemic in China has been largely controlled due to the implementation of many public health programs and policies targeting specific groups and geographical areas. Finding and supporting effective health programs will make it possible to achieve the World Health Organization's goal of controlling tuberculosis in China.

KEYWORDS

pulmonary tuberculosis, disease burden, spatial–temporal cluster analysis, prediction study, China

1 Introduction

Tuberculosis is an airborne contagious disease caused by the bacterium *Mycobacterium tuberculosis*. In recent years, TB has become the second deadliest contagious disease globally, surpassed only by COVID-19 (1). According to statistics, since 2019, the global new incidence rate of TB and the number of deaths due to TB have been rising, reversing the long-standing downward trend. In 2022, there were over 10 million new TB patients worldwide, and approximately 1.3 million people died from this disease. TB patients in eight countries account for more than two-thirds of the global total, including India, Indonesia, China, the Philippines, Pakistan and other countries (2). Consequently, TB remains one of the leading causes of adverse health effects and mortality worldwide, particularly in low- and middle-income developing countries.

As a developing country, China has prioritized TB prevention and control in alignment with the recommendations of the World Health Organization (3, 4). China has developed and implemented TB prevention and control programs such as the international widespread adoption of the directly observed treatment, short-course (DOTS) method across different provinces (5), etc. Despite these endeavors, China persists one of the high-burden countries for TB (1), encountering ongoing challenges in TB prevention and control, especially for the main type of pulmonary tuberculosis (PTB), accounting for about 78.7% of tuberculosis cases (6). Therefore, this research focuses on PTB specifically to analyze the Tuberculosis disease burden in China. Current research predominantly delves into the analysis of natural factors, including meteorological factors, seasonal patterns, and geographic locations, and their impact on TB transmission (7–9). Moreover, regional disparities in TB burden are also linked to social factors, including variations in TB-related prevention and control policies across different regions. Therefore, there is an urgent need for nationwide spatiotemporal cluster analysis to provide a basis for comparing the effectiveness of different policies in different regions, thereby pinpointing more targeted TB prevention and control measures. Such an approach will furnish scientific evidence to bolster the efficiency of TB prevention and control efforts, optimizing existing strategies.

Furthermore, predictive studies are pivotal in anticipating future trends in TB burden. Although geospatial predictive models have become increasingly utilized for this purpose (10), time series-based predictive models remain underexplored in research

in this realm. In mainland China, leveraging time-series prediction models rooted in historical data, researchers can forecast whether the forthcoming TB burden will align with the World Health Organization's end-tuberculosis milestones based on the current situation, offering scientific guidance for timely adjustments in planning and resource allocation.

In order to identify the factors causing disparities in TB prevention and control policies in different regions and to furnish theoretical underpinnings for the development of more efficacious and judicious policies and measures for TB prevention, this study aims to achieve several key objectives. Firstly, it seeks to delineate the current landscape of TB burden in China and examine variations in TB burden across different provinces. Additionally, it endeavors to forecast the future TB burden in China and evaluate the nation's progress toward meeting the targets outlined in the World Health Organization's End TB Strategy by 2030. To accomplish these aims, a time series forecast model is employed to project future trends in TB burden and provide scientific insights to facilitate timely adjustments in planning and resource allocation.

2 Materials and methods

2.1 Research data

In particular, PTB remains the predominant form of TB, constituting approximately 85% of all TB cases (1). Therefore, we conduct this study using PTB surveillance data. TB surveillance data on PTB pathogenicity testing and PTB disease burden for 31 provinces and cities in mainland China spanning from 2004 to 2018 were collected from the National Population and Health Scientific Data Sharing Platform - Public Health Scientific Data Center.¹ This comprehensive dataset encompasses 14,816,329 TB incidence cases and 42,465 TB deaths for analysis. By categorizing annual TB burden data across different age groups, we derived specific TB-related statistics for adolescents (0–19 years), young adults (20–39 years), middle-aged adults (40–59 years), and older adults (≥60 years).

1 <https://www.phsciencedata.cn/Share/index.jsp>

2.2 Statistical methods

The statistical methods used in this study encompass three distinct components which are the linear regression, cluster analysis, and projection prediction, respectively. The first part was to assess the burden of TB disease and discerning trends in the change in mainland China as well as across individual provinces. To accomplish this, we utilized the join-point regression model, a robust statistical tool for assessing trends in disease burdens over time (11). Join-point regression does not require the data to be strictly smooth and automatically detects trend change points in the data, which makes it suitable for public health data with non-linear trends. Additionally, the join-point regression model can automatically detect change points in data trends, providing more precise statistical analysis crucial for interpreting and predicting time trends in health-related events (12). We used annual percentage changes (APCs) (Equations 1-1, 1-2), and average annual percentage changes (AAPCs) (Equation 2-1) to assess trends in the morbidity and mortality of TB between 2004 and 2018. APCs were used to calculate segment-specific linear trends throughout the studied year. AAPCs were used to estimate overall changes throughout the entire duration of the study. These calculations were performed using the join-point model according to the following formula (13):

(1) APCs

$$\ln(\text{morbidity or mortality}) = \alpha + \beta x + \varepsilon \tag{1-1}$$

$$APCs = 100 \times (\exp(\beta) - 1) \tag{1-2}$$

(2) AAPCs

$$AAPCs = \left\{ \exp\left(\frac{\sum w_i \beta_i}{\sum w_i}\right) - 1 \right\} \times 100 \tag{2-1}$$

Where x represents the calendar year. α denotes the constant of the linear fitting; β is expressed as the slope coefficient of each year; ε is the residual error between the estimated and actual values (Equation 1-1). w_i is the number of PTB data for each year (Equation 2-1). To assess whether the fluctuation trend in different segments was statistically significant, we compared the AAPC to 0. If the lower limit of the 95% confidence interval (CI) of the AAPC exceeded zero, it indicated an increasing trend in morbidity or mortality. Conversely, if the upper limit of the 95% CI of the AAPC fell below zero, it suggested a decreasing trend in morbidity or mortality.

In the second part, we performed a systematic cluster analysis based on class averaging with Dynamic Time Wrapping (DTW) distance (Equations 3-1–3-3) on the time series of incidence and mortality in 31 provinces and cities in China. This analysis aimed to elucidate the diverse patterns of TB incidence and mortality control across different provinces. While Euclidean distance and DTW distance are commonly used matrix distance calculation methods in clustering algorithms (14), Euclidean distance is notably sensitive to small variations in the time axis. Even minor deviations in time series data may result in substantial differences in calculated distances, rendering it less suitable for time series data with varying time lengths. Moreover, the requirement for

equal time series lengths presents a significant limitation, necessitating padding or truncation of the time length, which may lead to inaccurate clustering and an inability to effectively handle non-linear changes (15). In contrast, DTW distance mitigates the shortcomings of Euclidean distance by accommodating time shifts and distortions. It aligns sequences by stretching or compressing them along the time axis, thereby offering flexibility to handle temporal offsets and enabling more precise measurements of similarities between time series. Unlike Euclidean distance, DTW distance does not mandate that time series be of the same length, nor does it necessitate preprocessing steps such as padding or truncation of the time axis. Moreover, DTW distance can effectively manage nonlinear distortions through multiple points (16). Given the significant disparities in morbidity and mortality rates among different provinces and time periods according to the PTB surveillance data, DTW distance was selected for utilization in the clustering process, which proceeds as outlined below:

(3) DTW

$$Q = q_1, q_2, q_3, \dots, q_i, \dots, q_m \tag{3-1}$$

$$C = c_1, c_2, c_3, \dots, c_j, \dots, c_n \tag{3-2}$$

$$DTW(Q, C) = \min \left\{ \frac{\sqrt{\sum_{k=1}^K \omega_k^2}}{K} \right\} \tag{3-3}$$

(4) Systematic clustering by class averaging

$$D_{PQ} = \frac{1}{n_p n_q} \sum_{x_i \in G_p} \sum_{x_j \in G_q} d_{ij}^2 \tag{4-1}$$

In the DTW distance calculation formula, (Equations 3-1–3-3) Q and C represent any two time series with respective lengths m and n , ω_k is the distance $DTW(q_i, c_j)$ between the corresponding points (q_i, c_j) of the two time series, K is the length of the longer time series. Following the computation of the DTW distance for each time series, we employ the class averaging method to systematically cluster each time series into distinct classes based on the calculated distances. Subsequently, we compute the distance between the new class and other classes using the same method. This iterative process continues as we merge the two closest classes into a new class according to the calculated distance. These steps are repeated until the number of classes approaches 1. At each stage, clustering diagrams are generated to aid in determining the optimal number of classes. In the systematic clustering model (Equation 4-1), the distance between classes is defined as the average of the squared distances between two elements within each class. Assuming that the clustering proceeds to a certain step where G_p and G_q are merged into a single entity. D_{PQ} represents the distance between G_p and G_q . Additionally, n_p and n_q are the number of elements contained within classes G_p and G_q respectively, while d_{ij} denotes the distance between two elements in G_p and G_q (Equation 4-1).

In the third part, we applied a projection model to forecast future TB incidence and mortality rates in China. Firstly, five prevalent time

series prediction models, ARIMA, PROPHET, GLMNET, RANDOMFOREST, and PROPHET BOOST were individually used (17–19); subsequently, we evaluated the prediction accuracy for both incidence and mortality rates, and found that the ARIMA model exhibited superior predictive efficiency for TB incidence and mortality time series data in China. In addition, the ARIMA model (Equation 5-1) has the characteristics of being easy to interpret and apply, being able to deal with nonlinear and unstable time series, and also being able to take into account the influence of historical data, with high accuracy in predicting the future, which is suitable for this retrospective study. Consequently, the ARIMA model was selected for subsequent predictions, and its formulation is as follows:

$$ARIMA(p,d,q)(P,D,Q)[m] \quad (5-1)$$

Where, (p,d,q) are the non-seasonal parameters in the model and $(P,D,Q)[m]$ represent the seasonal parameters in the model. p and P indicate the number of lags of the time series data itself (lags) used in the forecasting model, which is also known as AR/Auto-Regressive term; d and D represent the number of orders of differencing required for the time series data to achieve stability, which is also called as the Integrated term; q and Q represent the number of lags of the forecasting error (lags) used in the forecasting model, additionally referred to as the MA/Moving Average term.

3 Results

3.1 Overall tuberculosis disease burden and changing trends in mainland China in 2004–2018

Figure 1 illustrates the yearly fluctuations in new TB cases and TB-related deaths across mainland China from 2004 to 2018. The trend in the number of new cases per year closely mirrors that of the annual incidence rate, both reaching a peak in 2005 at 1,259,308 cases and 96.88 per 100,000 population, respectively, before experiencing a notable decline (Figure 1A). By 2018, the annual incidence rate had reduced to only 61.18% of its 2005 level. Similarly, the trend observed in the number of deaths aligns with that of the annual mortality rate, with both reaching their zenith in 2009 at 3,783 deaths and 0.28 per 100,000 population, respectively. However, following a six-year decline from 2009 to 2014, both indicators began to ascend in 2015 (Figure 1B). Notably, in 2018, the number of deaths and the annual mortality rate stood at 3,149 and 0.23 per 100,000 population, respectively, closely resembling the peak values observed in 2009. In summary, mainland China experienced an overall decrease in incidence percentage change of -20.59% from 2004 to 2018, while witnessing an increasing trend in mortality percentage change, which reached 105.35% (Supplementary Table S1).

Figure 2 depicts the distribution of new TB cases and deaths across various age groups in mainland China over the past 15 years, revealing significant age-related disparities. Notably, the burden of TB is lowest in the 0–19 age group, where the annual number of new cases remains below 105,000 and deaths do not exceed 30 per 100,000 annually, both exhibiting a gradual decline. In 2018, the number of new cases in this age group was 54,977, with 44 deaths reported. Between 2004 and 2011, the 20–39 age group exhibited the highest incidence rate among the four age groups, peaking at over 400,000 cases in 2005, though it steadily declined

thereafter. Subsequently, the 40–59 age group (from 2002 to 2015) and the age group over 60 years (from 2016 onwards) successively emerged as the primary cohorts for new TB cases, each surpassing 300,000 cases annually. Regarding mortality, the proportionate distribution among the four age groups has remained relatively stable over the years. Specifically, the age group 60 and above consistently exhibits the highest mortality rate, averaging over 1,500 cases per year, while the 40–59 and 20–39 age groups fall within the intermediate range, each averaging around 500 and 300 cases per year, respectively. In summary, the TB burden in mainland China predominantly afflicts individuals aged 60 years and above, markedly surpassing that observed in the 0–19 age group.

3.2 Comparison of TB disease burden and changing trends in 31 provinces and cities in mainland China in 2004–2018

Figure 3 presents the TB burden distribution among provinces in mainland China, revealing notable geographic disparities in the percentage change of TB burden from 2004 to 2018. Among them, Tibet (124.24%) and Xinjiang (114.72%) exhibited the highest percentage change in TB incidence, displaying a pronounced upward trend, whereas Zhejiang Province (-50.45%) and Jiangsu Province (-51.33%) demonstrated the smallest percentage change in TB incidence, reflecting a prominent downward trend compared to 2004 data (Supplementary Table S1). Significant percentage changes in the mortality rate of pulmonary TB were observed across various regions of mainland China. Notably, the four western regions (Xinjiang 1336.40% , Tibet 364.77% , Qinghai 328.31% , Chongqing 280.20%), six central regions (Anhui 379.22% , Hunan 272.28% , Hubei 196.15% , Henan 124.51% , Inner Mongolia 120.01% , Jiangxi 113.15%), and five eastern regions (Shandong 349.44% , Liaoning 320.85% , Hebei 171.26% , Hainan 157.27% , Guangxi 110.17%) displayed substantial percentage changes in mortality rates. Conversely, Tianjin (-52.25%) and Shanghai (-68.30%) exhibited relatively minor percentage changes in mortality rates. (Notably, due to a lack of mortality data in Ningxia in 2004, its percentage change value was excluded from the statistical calculations, Supplementary Table S1).

From 2004 to 2018, the APCs for TB incidence and mortality in the 31 provinces and cities indicated a statistically significant mean APC and 95% CI of -2.1 ($-3.2, -1.0$) for incidence (Supplementary Table S2 for specific results), while the AAPC of 4.1 ($-1.2, 9.8$) was not statistically significant (Supplementary Figures S1, S2). Notably, among the 31 provincial and municipal areas, Jilin, Jiangsu, Zhejiang, and Fujian provinces exhibited significant and statistically significant decreasing trends in morbidity, with AAPCs of -5.1 ($-9.1, -0.9$), -5.4 ($-6.3, -4.6$), -5.0 ($-6.1, -3.9$), and -5.0 ($-9.0, -0.9$), respectively (Supplementary Figure S1). Conversely, Xinjiang displayed the most noticeable increasing trend in mortality among the 31 provincial and municipal regions, with an AAPC of 17.7 ($7.3, 29$) (Supplementary Figure S2).

3.3 Time series clustering analysis of the burden of tuberculosis disease in mainland China

The time series analysis of TB disease burden in mainland China was clustered using DTW distance as the clustering index, and the

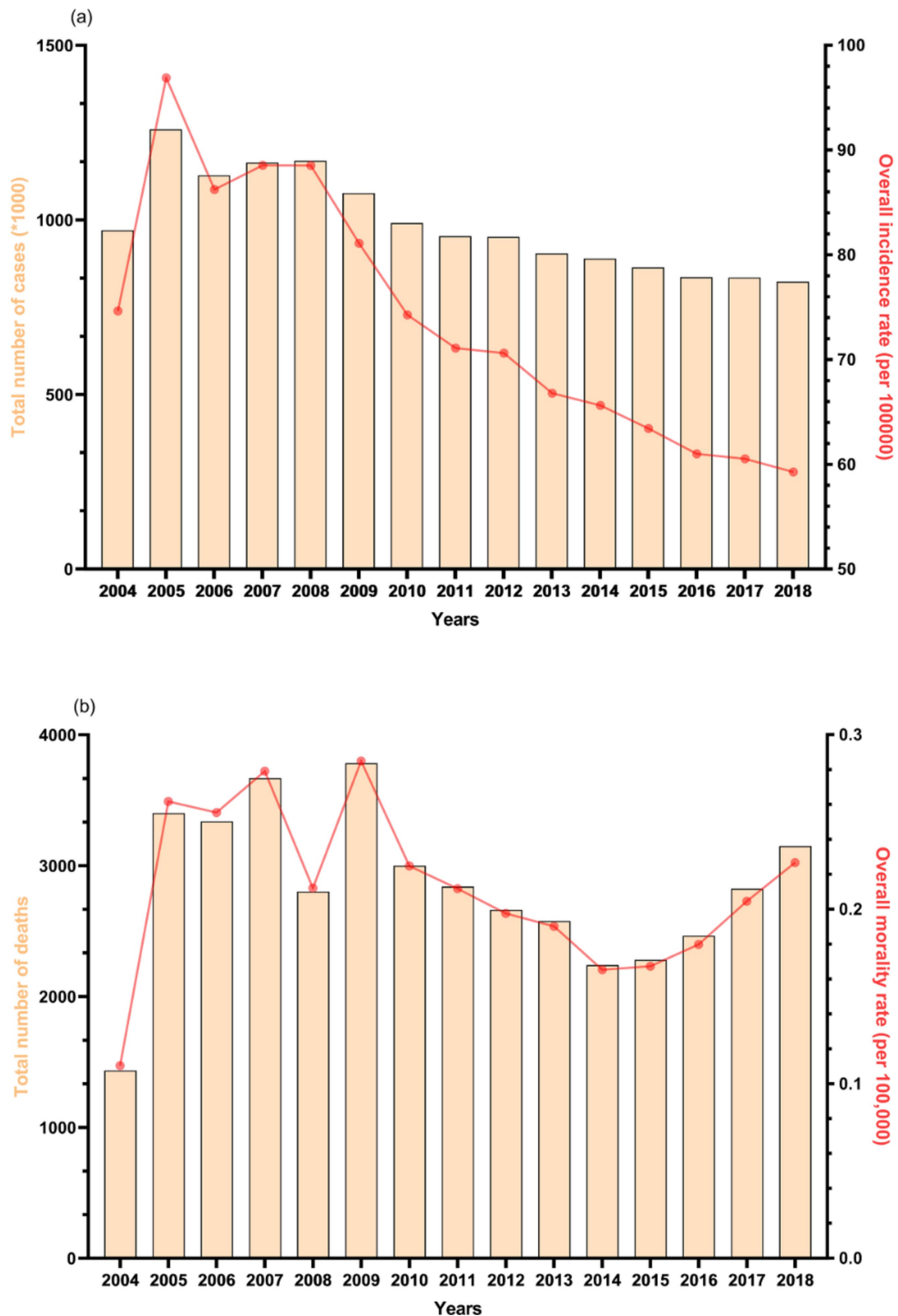


FIGURE 1 Total situations of (A) tuberculosis incidence and (B) tuberculosis deaths in mainland China, 2004–2018.

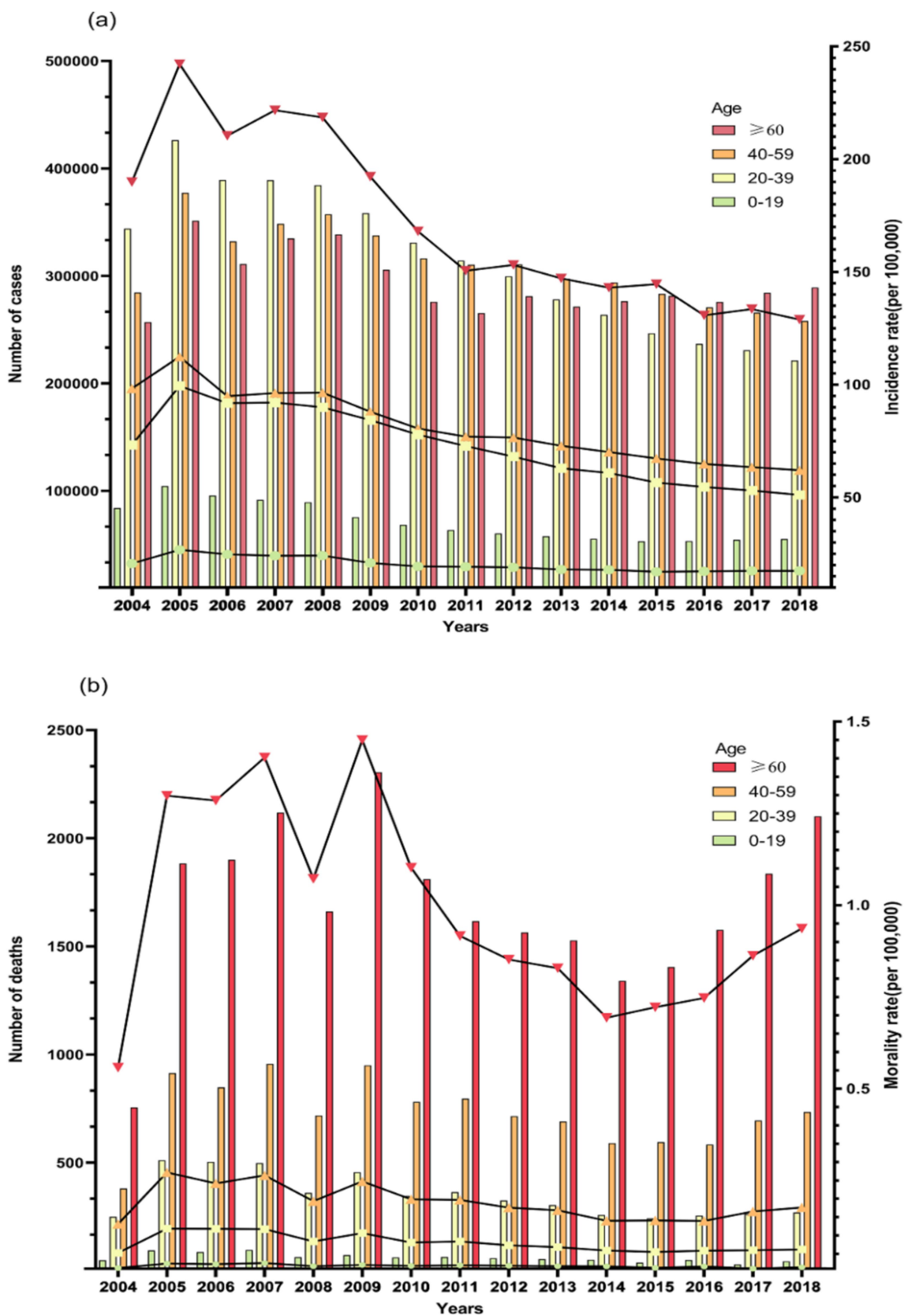
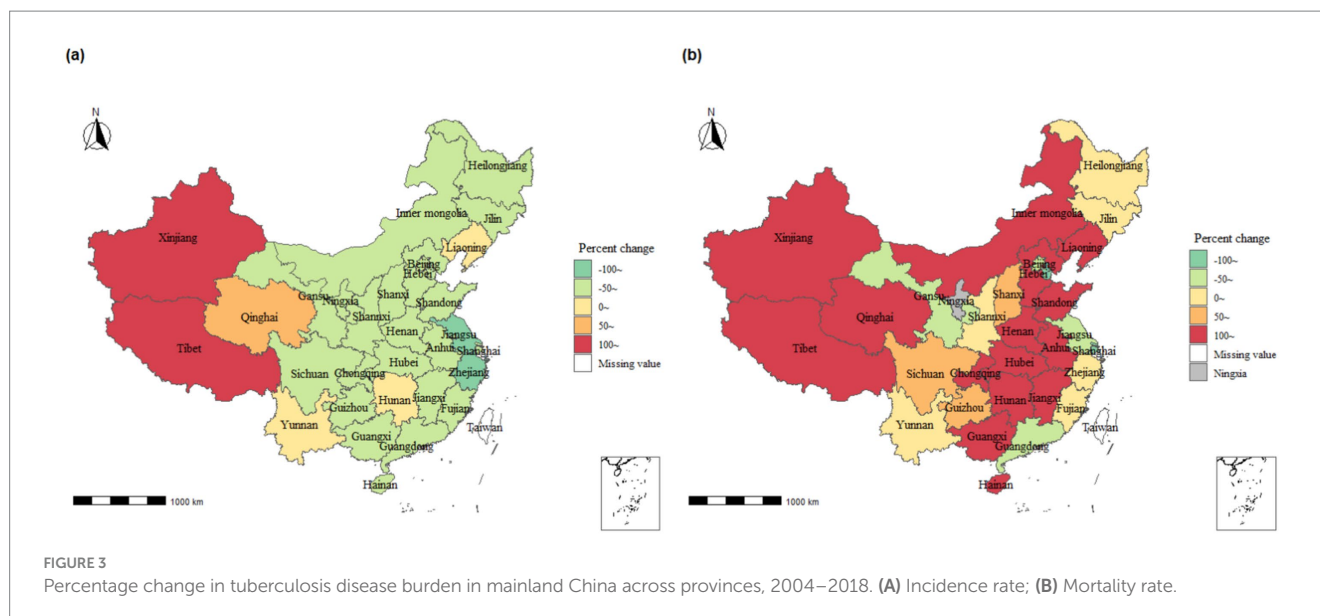


FIGURE 2 Situations of (A) tuberculosis incidence and (B) tuberculosis deaths by different age groups in mainland China, 2004–2018.



results were presented in Figure 4. The clustering results for TB incidence rates across provinces were categorized into seven major clusters, while the mortality rates were divided into six major clusters (Supplementary Figures S3–S15).

Cluster (I) represents the clustering results for morbidity:

- Categories 1, 3, and 5 show a continuous upward trend.
- Category 2 fluctuates between 120 and 190 and exhibits a single-peaked change.
- Categories 4, 6, and 7 show a continuous downward trend.

Cluster (M) represents the clustering results for mortality:

- Categories 1 and 4 exhibit an upward fluctuating trend.
- Categories 2 and 5 exhibit a downward fluctuating trend.
- Category 3 fluctuates between 0.2 and 0.7 and shows a “U” shaped fluctuation with multiple peaks.
- Category 6 fluctuates between 0.05 and 0.5 with multiple peaks, showing no clear upward or downward trend.

The time series analysis of TB incidence rates in most regions of mainland China, primarily the central and eastern regions, falls into Cluster (I)4. The incidence rates in this cluster gradually reduce within a single peak range of 100 to 120 per 100,000 (Supplementary Figure S6). In contrast, the time series analysis of TB mortality in most regions falls into Cluster (M)6, where mortality rates fluctuate within 0–0.5 per 100,000 with multiple peaks (Supplementary Figure S15).

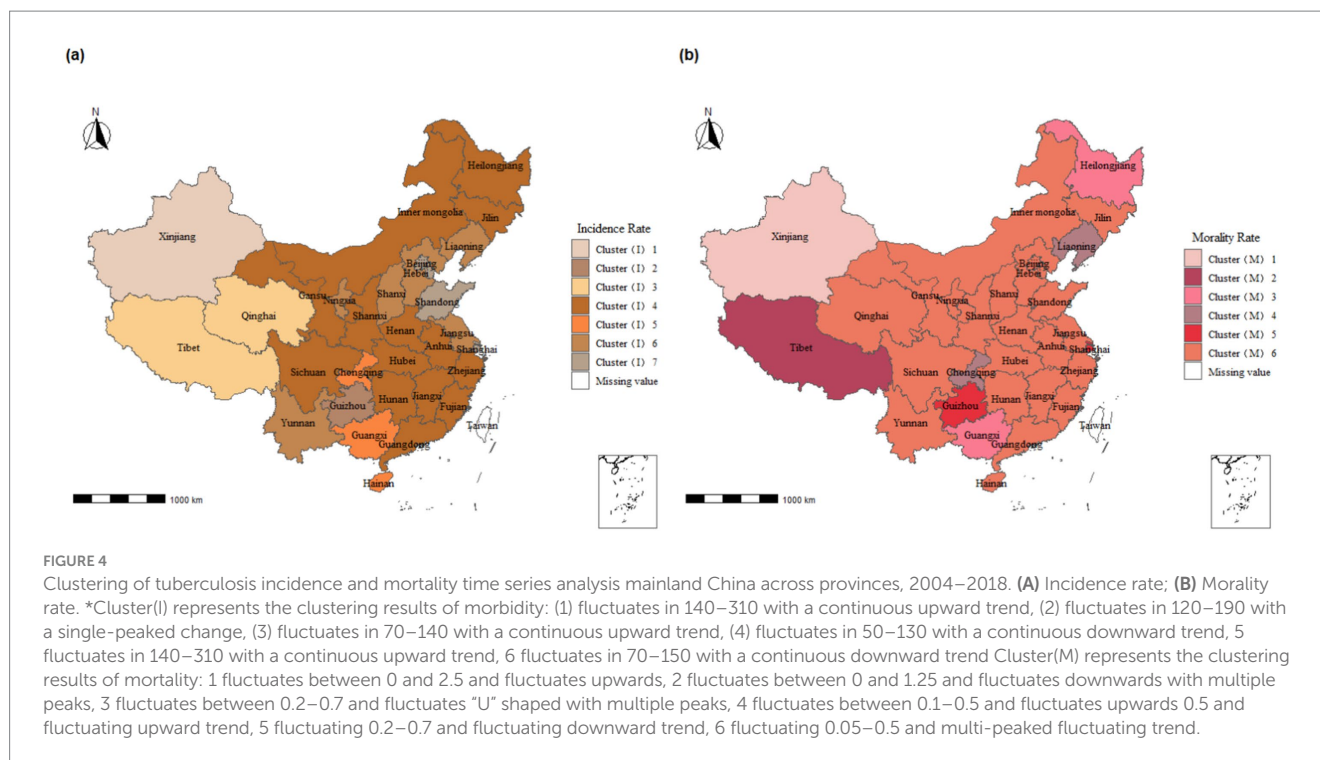
Specifically, we found that the time series clustering of morbidity in Tibet and Qinghai province belongs to Cluster (I)3, while the mortality clustering in these regions belongs to Cluster (M)2 and Cluster (M)6, respectively. The mortality clustering for Guizhou and Shanghai falls into Cluster (M)5, whereas the incidence rate clustering for Guizhou belongs to Cluster (I)2, and Shanghai belongs to Cluster (I)7. Notably, Xinjiang forms a unique cluster, distinct from all other provinces in both morbidity and mortality.

3.4 Model fitting and prediction for time series analysis of tuberculosis disease burden in mainland China in 2030

We found the ARIMA model was the most optimal prediction model for forecasting the TB disease burden in mainland China after testing of the five included models (Supplementary Tables S3, S4). This conclusion was drawn because the ARIMA model demonstrated the best prediction accuracy across several metrics, including MAE, MAPE, MASE, sMAPE, RMSE, and R^2 . The final prediction results suggested that the overall incidence rate would continue to decline by 2030, while the mortality rate would continue to increase by 2030 (Figure 5). In addition, the time series analysis of TB disease burden showed an obvious annual cycle and seasonality, with a trough in winter and a peak in spring (Figure 5; Supplementary Figures S16, S17).

4 Discussion

In this study, we utilized publicly available data on TB in mainland China provided by the Public Health Science Data Center. We analyzed the spatiotemporal distribution characteristics of TB cases from 2004 to 2018 and conducted the first-ever nationwide predictive research on the burden of TB in China. Regarding tuberculosis burden surveys, previous studies, both domestic and international, primarily focused on descriptive statistical analysis, often limiting their scope to analyzing population prevalence trends. This study goes beyond merely describing changes in disease burden by incorporating cluster analysis based on varying trend patterns. This approach allows for a more nuanced observation of the clustering characteristics of TB burden across regions, facilitating the identification of effective prevention and control strategies. Additionally, the inclusion of predictive analyses enhances the ability to evaluate the effectiveness of existing TB health policies and forecast future trends. The findings of this study indicate that the TB prevention and control policies in Guizhou Province and Xinjiang are



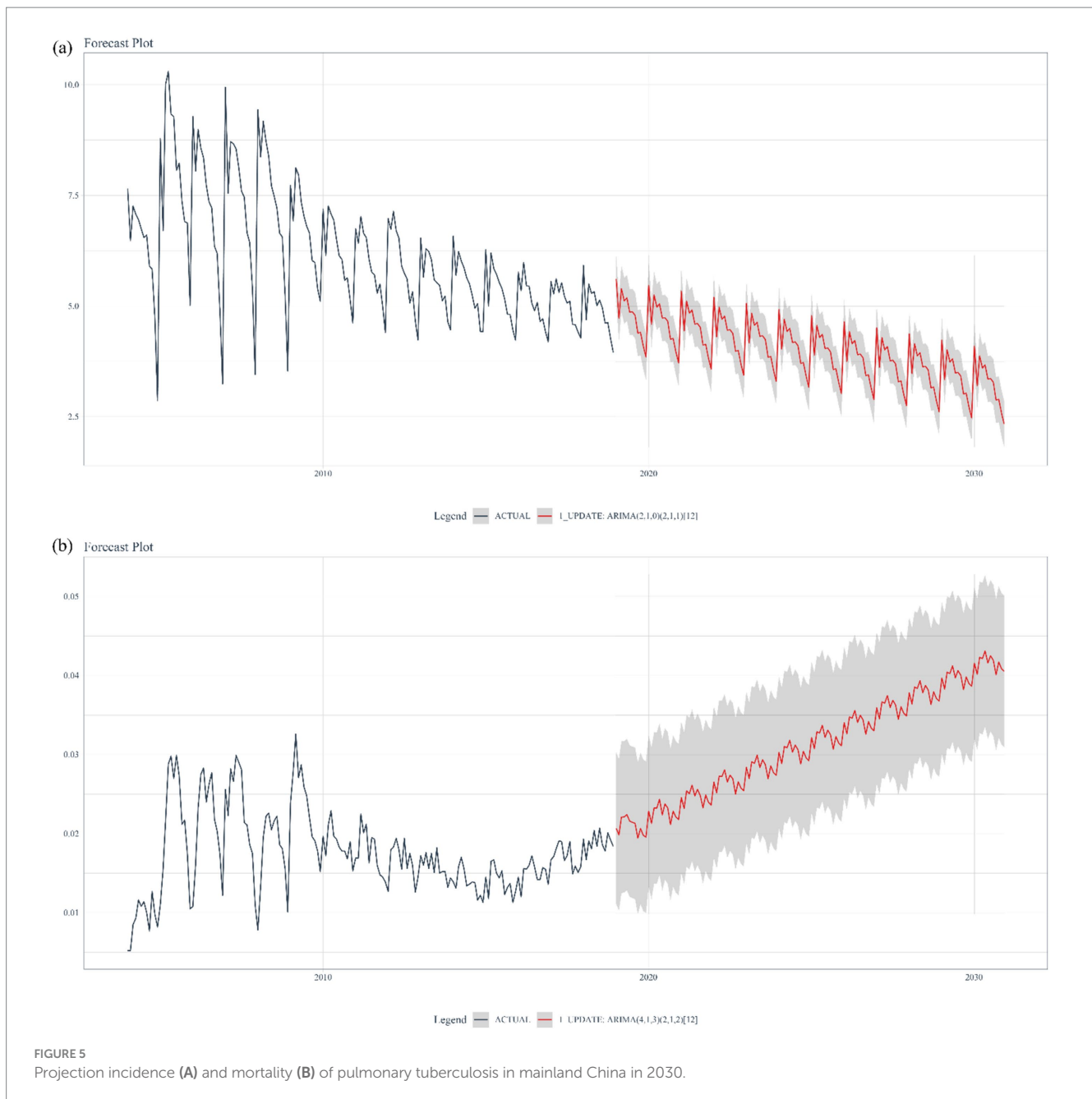
particularly effective and may serve as models for nationwide implementation.

We observed a sharp increase in reported TB cases during 2004–2005. This may be attributed to the launch of the National Tuberculosis Information Management System (TBIMS) by the Chinese Center for Disease Control and Prevention (CDC) in 2005. This system enabled the online collection of information on all TB cases nationwide (20).

The TB mortality rate, although peaking in 2005, has been on the rise since 2015 (Figure 1B). In response, many relevant government departments in China conducted a joint investigation and released an action plan (2019–2022) (21), formulating comprehensive action goals and key prevention and control measures at the national level. However, predictive results indicate that China may still face challenges in achieving the World Health Organization’s 2030 tuberculosis control objectives. This disparity is due to two main factors: (1) poor control of drug-resistant TB in underdeveloped regions due to economic imbalances (22, 23), and (2) an uneven distribution of disease risk caused by differences in TB-related policies at the provincial level in China (24). In light of this situation, China needs to rely on the findings of cluster and predictive research, compare specific policies among provinces, and explore more effective and rational TB prevention strategies to achieve the goals set by the World Health Organization’s End TB Strategy.

Furthermore, we observed a significantly higher burden of PTB among older adult individuals (≥ 60 years old) in China compared to the younger population. Similar findings have been reported in several provincial and municipal areas across the country (25–28). This phenomenon can be attributed to two main factors. Firstly, policy measures play a crucial role. The Chinese government has led initiatives to implement health education programs on TB among children and adolescents in schools, maintain school environmental hygiene, and improve timely reporting mechanisms for TB cases (29). These measures effectively reduce the incidence of TB among the

younger population. Secondly, the older adult population is more susceptible to TB due to their higher prevalence of underlying health conditions and weakened immune systems. This vulnerability increases their susceptibility to TB infection and raises the risk of reinfection (30–32). Current tuberculosis prevention and treatment policies should prioritize the older adult population by effectively raising TB awareness through educational outreach. Efforts should include BCG vaccine supplementation and the promotion of healthy lifestyles to boost immunity, ultimately helping to prevent TB transmission among older adults. Based on the results of cluster and time-series analyses, we observed regional characteristics in the spatiotemporal distribution of TB cases in China. Guizhou province, situated in mountainous areas, significantly lags behind Shanghai in terms of economic development. However, both Guizhou and Shanghai fall under Cluster (M) 5 regarding the trend of PTB mortality rates, showing a fluctuating downward trend within the range of 0.2/100,000 to 0.8/100,000. This fluctuating downward trend aligns with the global, European Region, and African Region trends reported in the Global Tuberculosis Report 2023 (1). Current research highlights that economic development levels and policy differences are crucial factors influencing the burden of PTB (33, 34). This suggests that the China Global Fund Tuberculosis Control Project, initiated and led by the Guizhou provincial government in 2003, played a pivotal role in controlling the spread of PTB within the province (35). The project proposes nine prevention and control measures to address the current situation and priorities of drug-resistant tuberculosis prevention and control in Guizhou Province. These measures encompass targeted strategies across the entire disease process, including screening, diagnosis, treatment, and prevention of drug-resistant TB. Additionally, practical preventive actions are recommended from the perspectives of patients, healthcare workers, and research experts. All of these efforts contribute to a more targeted approach to preventing and controlling the transmission of



drug-resistant TB. Additionally, the time-series clustering of incidence rates in Tibet and Qinghai Province placed them in the same Cluster (I)3, indicating a continuous upward trend. This may be due to the fact that the residents in these regions are primarily herders with relatively low literacy levels and limited economic means, preventing them from accessing local healthcare services. This situation leads to the continued spread of PTB, impacting its prevention and control efforts (34). Tibet and Qinghai can draw on Guizhou Province's TB prevention and control policies and update existing TB prevention and control measures by taking into account the specific characteristics of their populations. Local health departments should focus on the older adult population. Overall, the level of economic development significantly influences the clustering of TB morbidity and mortality. This influence stems from the direct impact of regional economic conditions on local income levels, healthcare services, and living

conditions, which in turn causes differences in the number of TB diagnoses and cures. Particularly in economically underdeveloped regions, the prevalence of TB is exacerbated by a higher concentration of individuals with limited education. These individuals often lack awareness about tuberculosis and are less likely to seek treatment in more developed areas. Additionally, the local medical infrastructure in these regions is often inadequate, failing to effectively control the spread of the disease, thereby increasing TB incidence. Furthermore, people in economically disadvantaged areas are more likely to be in close contact with live animals and livestock, a known risk factor for TB transmission. This factor further contributes to the higher concentration of TB cases in these economically less developed regions to some extent.

From 2004 to 2018, the number of PTB cases in Xinjiang showed a gradual increase and was significantly higher than the national

average. One possible reason for this trend is that Xinjiang has implemented a series of measures to screen for cases and control PTB outbreaks. By the end of 2010, Xinjiang's DOTS coverage had reached 100%, and in 2013 it was the first province in the country to implement a new "trinity" model of TB control services. This model includes policies such as "centralized isolation treatment for infected patients and home treatment for non-infected patients, such as centralized medication + nutritious breakfast" (24). Although the number of TB cases appears to be increasing, this is actually a reflection of the continuous improvement in the TB reporting system, which should be maintained. Lessons can be drawn from Guizhou Province, which not only implements effective health policies within the province for TB prevention and control but also actively learns from more developed regions. For instance, Guizhou has invited experts from cities like Shanghai to provide scientific lectures and has adopted successful prevention and control strategies from Shanghai, such as enhancing the TB reporting system and ensuring timely follow-up for individuals infected with tuberculosis bacilli. National epidemiological surveys on tuberculosis should be conducted regularly to identify provinces with notable success in TB prevention and control. Promoting these best practices across provinces, increasing opportunities for interprovincial learning and exchange, and improving collaboration will ultimately enhance the overall efficiency of TB prevention and control efforts.

Using ARIMA model projections, we found an upward trend in TB mortality in mainland China since 2015. This trend is influenced by several factors. Firstly, there has been an increase in the burden of drug-resistant TB, with approximately 450,000 new cases of rifampicin-resistant TB reported in 2021 (36). Secondly, the impact of the COVID-19 pandemic led to the suspension of TB case reporting in 2021, potentially setting back TB control efforts by 8 years in achieving the goal of ending TB by 2030 (37). Pandemics have been shown to impact TB case reporting and treatment outcomes. Both TB and COVID-19 tests are more likely to yield false-negative results in a TB pandemic setting (38). Both diseases require better samples with independence to improve diagnostic precision and accuracy. Airborne and aerosol-based transmission are the primary modes of spread for both SARS-CoV-2 and *Mycobacterium tuberculosis*. Additionally, COVID-19 worsen the TB disease burden by increasing pulmonary disability and lung damage (39). To enhance positive case identification, numerous studies have conducted geospatial analyses to delineate the distribution of TB cases (40, 41). While TB is frequently prevalent in low-income areas and regions with inadequate sanitation, higher disease prevalence is often observed among individuals without stable residency status, including refugees, asylum seekers, and ordinary migrants (42). Infectious disease surveillance among incoming mobile populations should be enhanced, with consistent registration, screening, and testing for tuberculosis infection. The implementation of closure control policies due to the pandemic may have added complexity to the actual TB disease burden in mainland China, necessitating further investigation.

This research is subject to several potential limitations related to the dataset and clustering algorithms. Firstly, while the National Population and Health Scientific Data Sharing Platform provides high-quality data, it lacks comprehensive personal information about patients. This limitation prevents the analysis of certain demographic characteristics, such as gender and occupation, which are significant factors influencing TB distribution. Additionally, because the

platform's data is not updated to the most recent year, only retrospective studies can be conducted, introducing potential selection bias. Limitations in the inclusion criteria for new cases further constrain the study. On the algorithmic side, the DTW algorithm also presents limitations. It requires substantial computational resources, particularly when handling large datasets, and it focuses solely on the similarity between two sequences, without incorporating actual geographic information (16). These factors contribute to the study's limitations. Firstly, PTB cases reported through the direct network reporting of infectious diseases may be subject to data bias due to variations in TB diagnostic capacity and healthcare provider practices. Secondly, our analysis was conducted at the provincial level, which may result in some level of data aggregation or disaggregation due to geographical scale effects. Thirdly, due to constraints in data collection, we were unable to explore in depth the etiological relationship between TB and risk confounding variables such as socioeconomic and environmental factors through multivariate analyses. However, despite these limitations, this study underscores the critical importance of PTB regulation. In the future, the prevention and treatment of tuberculosis will usher in a new phase of challenges in the post-epidemic era, necessitating both traditional, proven methods such as vaccination and the timely adaptation of prevention and treatment policies to more effectively achieve the World Health Organization's (WHO) goal of eliminating tuberculosis. To support these objectives, research should prioritize two main areas: first, evaluating intervention effectiveness through health economics to identify the most impactful strategies, and second, leveraging advancements in artificial intelligence to develop efficient surveillance tools that can pinpoint high-risk areas and populations. This approach will enable a more responsive infectious disease system at all levels, facilitating the rapid detection of potential infection sources and the swift implementation of preventive and control measures to reduce transmission risks.

5 Conclusion

While China's TB incidence continues to decline steadily, the burden remains significant, particularly due to regional inequalities and a concentration of cases among the older adult. These findings underscore the urgent need to enhance protective and preventive measures, along with intensified TB-related awareness campaigns and educational initiatives, to align with the World Health Organization's milestones for TB control in China.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary material](#).

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the

patients/participants OR patients/participants legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

JG: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. CL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. FL: Data curation, Formal analysis, Investigation, Validation, Visualization, Writing – original draft. EZ: Data curation, Formal analysis, Investigation, Validation, Visualization, Writing – original draft. RM: Data curation, Formal analysis, Investigation, Validation, Visualization, Writing – original draft, Methodology. LZ: Data curation, Formal analysis, Investigation, Validation, Visualization, Writing – original draft. BL: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors express their sincere gratitude to the Chinese Centers for Disease and Prevention (CDC) for organizing the creation of the

References

1. World Health Organization. Global tuberculosis report 2023 (2023). Available at: <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023> [Accessed Apr 23, 2024].
2. Espindola AL, Varughese M, Laskowski M, Shoukat A, Heffernan JM, Moghadas SM. Strategies for halting the rise of multidrug resistant TB epidemics: assessing the effect of early case detection and isolation. *Int Health*. (2017) 9:80–90. doi: 10.1093/inthealth/ihw059
3. Xinhua NET. Peng Liyuan calls for increased input, cooperation to advance global TB control. (2023). Available at: <https://english.news.cn/20230323/6c66cd31f2024a979d1d85de28f4b1d7/c.html> [Accessed May 22, 2023].
4. State Council Office of the People's Republic of China. Circular of the general Office of the State Council on the issuance of the National Tuberculosis Prevention and control plan for the thirteenth five-year plan (2017). Available at: https://www.gov.cn/zhengce/content/2017-02/16/content_5168491.htm [Accessed May 23, 2023].
5. Wang L, Zhang H, Ruan Y, Chin DP, Xia Y, Cheng S, et al. Tuberculosis prevalence in China, 1990–2010; a longitudinal analysis of national survey data. *Lancet*. (2014) 383:2057–64. doi: 10.1016/S0140-6736(13)62639-2
6. Wang X, Jiang M, Pang Y, Sun D, Yu C, Wang L, et al. Current status of tuberculosis burden in China. *Zhonghua Liu Xing Bing Xue Za Zhi*. (2024) 45:857–64. doi: 10.3760/cma.j.cn112338-20240311-00111
7. Li Q, Liu M, Zhang Y, Wu S, Yang Y, Liu Y, et al. The spatio-temporal analysis of the incidence of tuberculosis and the associated factors in mainland China, 2009–2015. *Infect Genet Evol*. (2019) 75:103949. doi: 10.1016/j.meegid.2019.103949
8. Li XX, Wang LX, Zhang H, Du X, Jiang SW, Shen T, et al. Seasonal variations in notification of active tuberculosis cases in China, 2005–2012. *PLoS One*. (2013) 8:e68102. doi: 10.1371/journal.pone.0068102

Public Health Science Data Center, which provides excellent data resources for researchers.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

The views expressed in this article are those of the authors and do not represent any official policy. All the authors have declared no relationships or activities that could appear to have influenced this work.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2024.1436515/full#supplementary-material>

9. Li XX, Wang LX, Zhang J, Liu YX, Zhang H, Jiang SW, et al. Exploration of ecological factors related to the spatial heterogeneity of tuberculosis prevalence in P. R. *China Glob Health Action*. (2014) 7:23620. doi: 10.3402/gha.v7.23620
10. Dao TP, Hoang XHT, Nguyen DN, Huynh NQ, Pham TT, Nguyen DT, et al. A geospatial platform to support visualization, analysis, and prediction of tuberculosis notification in space and time. *Front Public Health*. (2022) 10:973362. doi: 10.3389/fpubh.2022.973362
11. Zhang Y, Liu J, Han X, Jiang H, Zhang L, Hu J, et al. Long-term trends in the burden of inflammatory bowel disease in China over three decades: a joinpoint regression and age-period-cohort analysis based on GBD 2019. *Front Public Health*. (2022) 10:994619. doi: 10.3389/fpubh.2022.994619
12. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med*. (2000) 19:335–51. doi: 10.1002/(SICI)1097-0258(20000215)19:3<335::AID-SIM336>3.0.CO;2-Z
13. Qiu P, He H, Zhao Y, Yang Z, Li S, Ni P, et al. Global disease burden attributed to unsafe sex in 204 countries and territories from 1990 to 2019: results from the global burden of disease study 2019. *Sci Rep*. (2023) 13:12900. doi: 10.1038/s41598-023-40001-2
14. Holder C, Middlehurst M, Bagnall A. A review and evaluation of elastic distance functions for time series clustering. *Knowl Inf Syst*. (2024) 66:765–809. doi: 10.1007/s10115-023-01952-0
15. Kapil S, Chawla M. Performance evaluation of K-means clustering algorithm with various distance metrics. 2016 IEEE Int Conf on Power Electronics, Intelligent Control and Energy Systems (ICPEICES) (2016) 1–4
16. Keogh E, Ratanamahatana C. Exact indexing of dynamic time warping. *Knowl Inf Syst*. (2005) 7:358–86. doi: 10.1007/s10115-004-0154-9
17. Sardar I, Akbar MA, Leiva V, Alsanad A, Mishra P. Machine learning and automatic ARIMA/Prophet models-based forecasting of COVID-19: methodology,

- evaluation, and case study in SAARC countries. *Stoch Environ Res Risk Assess.* (2023) 37:345–59. doi: 10.1007/s00477-022-02307-x
18. Ospina R, Gondim AMJ, Leiva V, Castro C. An overview of forecast analysis with ARIMA models during the COVID-19 pandemic: methodology and case study in Brazil. *Mathematics.* (2023) 11:3069. doi: 10.3390/math11143069
19. Morid AM, Sheng O, Dunbar J. Time series prediction using deep learning methods in healthcare. *ACM Trans Manage Info Syst.* (2023) 14:1–29. doi: 10.1145/3531326
20. Huang F, Cheng S, Du X, Chen W, Scano F, Falzon D, et al. Electronic recording and reporting system for tuberculosis in China: experience and opportunities. *J Am Med Inform Assoc.* (2014) 21:938–41. doi: 10.1136/amiajnl-2013-002001
21. Health Commission of the People's Republic of China, Development and Reform Commission, Ministry of Education, Ministry of Science and Technology, Ministry of Civil Affairs, Ministry of Finance, Office of Poverty Alleviation, Medical Insurance Bureau. Stop TB action plan (2019–2022) (2019). Available at: https://www.gov.cn/gongbao/content/2019/content_5437149.htm [Accessed May 27, 2023].
22. Lecai J, Mijiti P, Chuangyue H, Qian G, Weiguo T, Jihong C. Treatment outcomes of multidrug-resistant tuberculosis patients receiving ambulatory treatment in Shenzhen, China: a retrospective cohort study. *Front Public Health.* (2023) 11:1134938. doi: 10.3389/fpubh.2023.1134938
23. Zheng XB, Diwan VK, Zhao Q, Hu Y, Bruchfeld J, Jiang WL, et al. Treatment quality and outcome for multidrug-resistant tuberculosis patients in four regions of China: a cohort study. *Infect Dis Poverty.* (2020) 9:97. doi: 10.1186/s40249-020-00719-x
24. Hu MFY, Li T, Zhao Y, Wang J, Xu C, Chen W. Unbalanced risk of pulmonary tuberculosis in China at the subnational scale: spatiotemporal analysis. *JMIR Public Health Surveill.* (2022) 8:e36242. doi: 10.2196/36242
25. Huang L, Li XX, Abe EM, Xu L, Ruan Y, Cao CL, et al. Spatial-temporal analysis of pulmonary tuberculosis in the northeast of the Yunnan province, People's Republic of China. *Infect Dis Poverty.* (2017) 6:53. doi: 10.1186/s40249-017-0268-4
26. Yu Y, Wu B, Wu C, Wang Q, Hu D, Chen W. Spatial-temporal analysis of tuberculosis in Chongqing, China 2011–2018. *BMC Infect Dis.* (2020) 20:531. doi: 10.1186/s12879-020-05249-3
27. Ge E, Zhang X, Wang X, Wei X. Spatial and temporal analysis of tuberculosis in Zhejiang Province, China, 2009–2012. *Infect Dis Poverty.* (2016) 5:11. doi: 10.1186/s40249-016-0104-x
28. Zhang Q, Ding H, Gao S, Zhang S, Shen S, Chen X, et al. Spatiotemporal changes in pulmonary tuberculosis incidence in a low-epidemic area of China in 2005–2020: retrospective spatiotemporal analysis. *JMIR Public Health Surveill.* (2023) 9:e42425. doi: 10.2196/42425
29. Chinese Center for Disease Control and Prevention. Guidelines for tuberculosis prevention and control in Chinese schools (2020). Available at: <https://www.gov.cn/zhengce/zhengceku/2020-12/05/5567137/files/85d293d4f3e04bacb420f252c2189278.pdf> [Accessed Jun 22, 2023].
30. Gao L, Lu W, Bai L, Wang X, Xu J, Catanzaro A, et al. Latent tuberculosis infection in rural China: baseline results of a population-based, multicentre, prospective cohort study. *Lancet Infect Dis.* (2015) 15:310–9. doi: 10.1016/S1473-3099(14)71085-0
31. Gao L, Bai L, Liu J, Lu W, Wang X, Li X, et al. Annual risk of tuberculosis infection in rural China: a population-based prospective study. *Eur Respir J.* (2016) 48:168–78. doi: 10.1183/13993003.00235-2016
32. Lu P, Liu Q, Zhou Y, Martinez L, Kong W, Ding X, et al. Predictors of discordant tuberculin skin test and quantiFERON-TB gold in-tube results in eastern China: a population-based. *Cohort Study Clin Infect Dis.* (2021) 72:2006–15. doi: 10.1093/cid/cia519
33. Wang YS, Wang JM, Wang WB. Temporal-spatial distribution of tuberculosis in China, 2004–2016. *Zhonghua Liu Xing Bing Xue Za Zhi.* (2020) 41:526–31. doi: 10.3760/cma.j.cn112338-20190614-00441
34. Yang GWJ, Zhang B, Wang T. Analysis on epidemiological characteristics of tuberculosis in Tibet autonomous region from 2008 to 2017. *Chin J Dis Control Preven.* (2020) 24:1106–9. doi: 10.16462/j.cnki.zbjkz.2020.09.023
35. Chen W, Li Y, Yuan W, Li YH, Song QF, Ye L, et al. Evaluate the eleven-year implementation effect of global fund to fight tuberculosis in Guizhou province. *China Health Industry.* (2016) 13:3. doi: 10.16659/j.cnki.1672-5654.2016.10.038
36. Su W, Ruan YZ, Li T, Du X, Jiang JW, Li RZ. Characteristics of rifampicin-resistant tuberculosis detection in China, 2015–2019. *Infect Dis Poverty.* (2021) 10:99. doi: 10.1186/s40249-021-00883-8
37. Hogan AB, Jewell BL, Sherrard-Smith E, Vesga JF, Watson OJ, Whittaker C, et al. Potential impact of the COVID-19 pandemic on HIV, tuberculosis, and malaria in low-income and middle-income countries: a modelling study. *Lancet Glob Health.* (2020) 8:e1132–41. doi: 10.1016/S2214-109X(20)30288-6
38. Dheda K, Perumal T, Moultrie H, Perumal R, Esmail A, Scott AJ, et al. The intersecting pandemics of tuberculosis and COVID-19: population-level and patient-level impact, clinical presentation, and corrective interventions. *Lancet Respir Med.* (2022) 10:603–22. doi: 10.1016/S2213-2600(22)00092-3
39. Roure S, Vallès X, Sopena N, Benítez RM, Reynaga EA, Bracke C, et al. Disseminated tuberculosis and diagnosis delay during the COVID-19 era in a Western European country: a case series analysis. *Front Public Health.* (2023) 11:1175482. doi: 10.3389/fpubh.2023.1175482
40. Chihota VN, Niehaus A, Streicher EM, Wang X, Sampson SL, Mason P, et al. Geospatial distribution of *Mycobacterium tuberculosis* genotypes in Africa. *PLoS One.* (2018) 13:e0200632. doi: 10.1371/journal.pone.0200632
41. Gehlen M, Nicola MRC, Costa ERD, Cabral VK, de Quadros ELL, Chaves CO, et al. Geospatial intelligence and health analytics: its application and utility in a city with high tuberculosis incidence in Brazil. *J Infect Public Health.* (2019) 12:681–9. doi: 10.1016/j.jiph.2019.03.012
42. Arshad S, Bavan L, Gajari K, Paget SN, Baussano I. Active screening at entry for tuberculosis among new immigrants: a systematic review and meta-analysis. *Eur Respir J.* (2010) 35:1336–45. doi: 10.1183/09031936.00054709