

Factors affecting the incidence of pulmonary tuberculosis based on GTWR model in China, 2004-2021

Hairu Yu¹, Jiao Yang¹, Yexin Yan¹, Hui Zhang¹, Qiuyuan Chen¹, Liang Sun^{1*}

¹ College of Public Health, Zhengzhou University, Zhengzhou 450001, Peoples R China

Corresponding author: Liang Sun

E-mail: Doubleliang@126.com

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

Abstract Contra-posing panel data on the incidence of pulmonary tuberculosis (PTB) at provincial level in China during 2004-2021 and introducing Geographically and Temporally Weighted Regression model (GTWR) model were used to explore the effect of various factors on incidence of PTB from the perspective of spatial heterogeneity. The Principal Component Analysis (PCA) was used to extract the main information from twenty-two indexes under six macro factors. The main influencing factors were determined by Spear-man correlation and multi-col-linearity test. After fitting different models, GTWR model was used to analyze and obtain the distribution changes of regression coefficients. Six macro factors and incidence of PTB were both correlated and there were no col-linearity between the variables. The fitting effect of GTWR model was better than Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) model. The incidence of PTB in China was mainly affected by six macro factors, namely medicine and health, transportation, environment, economy, disease and educational quality. The influence degree showed an unbalanced trend in the spatial and temporal distribution.

Keywords: pulmonary tuberculosis; Geographically and Temporally Weighted Regression model; influencing factors;

Introduction

According to the Global Tuberculosis Report 2022[1] released by the World Health Organization (WHO), the number of 10.6 million new tuberculosis (TB) cases occurred worldwide in 2021, and China remained one of the countries with the highest burden of TB. In 2021, China had the third highest number of TB cases after India and Indonesia. The incidence of TB was affected by many factors, such as society-economy, population, climate, transportation and other factors, which making TB become a multi-faceted and complex public health problem[2, 3]. Existing studies analyzed and explored the influencing factors of TB onset in China[4, 5]. However, most of studies used OLS and GWR model to analyze influencing factors, and did not consider the time and space dimensions, which cannot accurately reflect the spatial-temporal heterogeneity of PTB onset and affecting factors in China. GTWR model was used to study the spatial-temporal heterogeneity of various diseases[6, 7, 8]. Therefore, to provide policies and measures for the regional prevention and treatment of PTB, we used the latest data of the national PTB incidence rate during 2004-2021, as well as twenty-two categories of indicators under the six macro-influencing factors, and analyzed the influencing factors of the PTB incidence of PTB based on GTWR model. so as to provide policies and measures for the regional prevention and treatment of PTB.

Methods

Research Data

Data on PTB incidence in China during 2004–2018 was collected from the Data Center of Public Health Science. and the data during 2019-2021 was collected from the China Health Statistics Yearbook. The influencing factors of medical and health, transportation, environment,

economy, disease and educational quality on the incidence of PTB in China were comprehensively considered. A total of twenty-two indicators under six macros (Table 1) were collected from 31 provinces, autonomous regions and municipalities (excluding Hong Kong, Macao and Taiwan) during 2004-2021. The vector map of China's provincial administrative divisions was downloaded by the National Center for Basic Geographic Information System.

Table 1. Measurement indicators of macro influencing factors

Macro influencing factor	Measure index
Medical and health care	Number of health workers (10000)
	Number of beds in medical and health institutions (10000)
	Number of medical and health institutions (per)
	Local expenditure on medical and health care (100 million yuan)
Transportation	Passenger turnover (100 million person-kilometers)
	Automobile ownership in highway operation (10000 units)
	Highway mileage (ten thousand kilometers)
	Passenger volume (10000)
Environment	Civil automobile ownership (10000 units)
	SO ₂ emissions (10000 tons)
	Smoke (powder) dust discharge (10000 tons)

	Ammonia nitrogen emissions (10000 tons)
	Chemical oxygen demand emissions (10000 tons)
	Per capita disposable income of all residents (Yuan)
Economy	PGDP (yuan)
	Per capita consumption expenditure of all households (Yuan)
	Consumer Price Index (last year =100)
Disease	AIDS incidence rate (per 10000)
	Number of regular institutions of higher Learning (institutions)
	Total number of staff and staff in regular institutions of higher learning (10000)
Educational quality	Average Number of Students in Institutions of Higher Learning per 100000 population
	Educational expenditure (10000 Yuan)

SO2= Sulphur dioxide; PGDP=Per Capital Gross Domestic Product; AIDS=Acquired Immune Deficiency Syndrome

Principal Component Analysis

Principal Component Analysis is a multivariate statistical method that converts multiple indicators under six macro factors into comprehensive indicators with little loss of information by using the method of dimensional reduction. Kaiser-Meyer-Olkin (*KMO*) and Bartlett tests

were used in this study. And all macro influencing factors were standardized and normalized.

Correlation analysis and col-linearity diagnosis

R 4.2.0 was used to conduct Spearman correlation test between each macro influencing factor and the incidence of PTB. And no correlation indicators were excluded. The standardized coefficient and Variance Inflation Factor (*VIF*) were calculated by linear regression. When $VIF \geq 10$, it indicated that there was a multi-col-linearity problem between variables. The multi-col-linearity index was removed and the optimal combination was obtained after several fitting experiments.

Geographically and Temporally Weighted Regression model

Traditional global regression models OLS cannot reflect the spatial heterogeneity of different regional coefficients, nor can they effectively excavate important local features between explanatory variables and explained quantities. GWR can only be used for cross-sectional data and could not consider time factor[9]. In order to consider the information of time and space, Huang Bo proposed GTWR model[10]. The GTWR model determines the shadow specific gravity of other sample points on the regression sample points by constructing the space-time weight matrix. Therefore, the space-time weight matrix plays a core role in the calculation process of the GTWR model. Its form is a diagonal matrix, and the elements in the matrix are determined by three factors, such as space bandwidth, kernel function and distance calculation formula. This paper was based on adaptive bandwidth, Gaussian kernel function and Euclidean distance, and determined by AICc criterion.

Results

Principle Component Analysis

The result of Principal component analysis was illustrated in Table 2 below (Supplementary material 1). Principal component analysis of *KMO* test values were greater than 0.06, $P < 0.05$.

Table 2. Normalized values for the principal component scores

NO.	Province	Time	Medicine and health	Transport ation	Environment	Econo my	Disease	Educational quality
1	Beijing	2004	0.084	0.074	0.091	0.234	0.0219	0.487
2	Tianjin	2004	0.037	0.012	0.094	0.148	0.0058	0.225
3	Hebei	2004	0.174	0.305	0.684	0.062	0.0014	0.272
4	Shanxi	2004	0.107	0.119	0.665	0.062	0.0048	0.178
...
555	Gansu	2021	0.266	0.172	0.136	0.297	0.093	0.277
556	Qinghai	2021	0.074	0.057	0.032	0.349	0.149	0.079
557	Ningxia	2021	0.058	0.044	0.055	0.375	0.068	0.136
558	Xinjiang	2021	0.258	0.242	0.246	0.36	0.253	0.280

Spear-man correlation test

Spear-man correlation test showed that six macro factors and incidence of PTB were correlated. The P values between medicine and health, transportation, environment, economy, disease and educational quality and the incidence of PTB are <0.001 , <0.001 , 0.004 , <0.001 , 0.003 and <0.001 , respectively.

Multi-col-linearity test

The result of multi-col-linearity test (Table 3) showed that each variable's *VIF* was less than 10, and the variable did not exist col-linearity problem.

Table3.The *VIF* index of macro factors of incidence of PTB

	Standardized coefficient	<i>VIF</i>
Medicine and health	-0.187	7.742
Transportation	0.358	7.211
Environment	-0.073	2.209
Economy	-0.297	2.62
Disease	0.208	1.314
Educational quality	-0.472	6.034

The comparisons of OLS, GWR, GTWR model

Based on fitting effect of different model, the result of comparisons was shown in the Table 4. The GTWR model had highest R^2 value and lowest $AICc$ value compared with other two models, demonstrating that the GTWR model outperformed OLS and GWR for determining the relations between incidence of PTB and six macro factors.

Table 4 .Values of R^2 and $AICc$ of OLS, GWR, GTWR model

	OLS	GWR	GTWR
R^2	0.400696	0.874037	0.88975
$AICc$	5342.206501	4576.42	4534.51

Regression coefficient characteristics of space and time

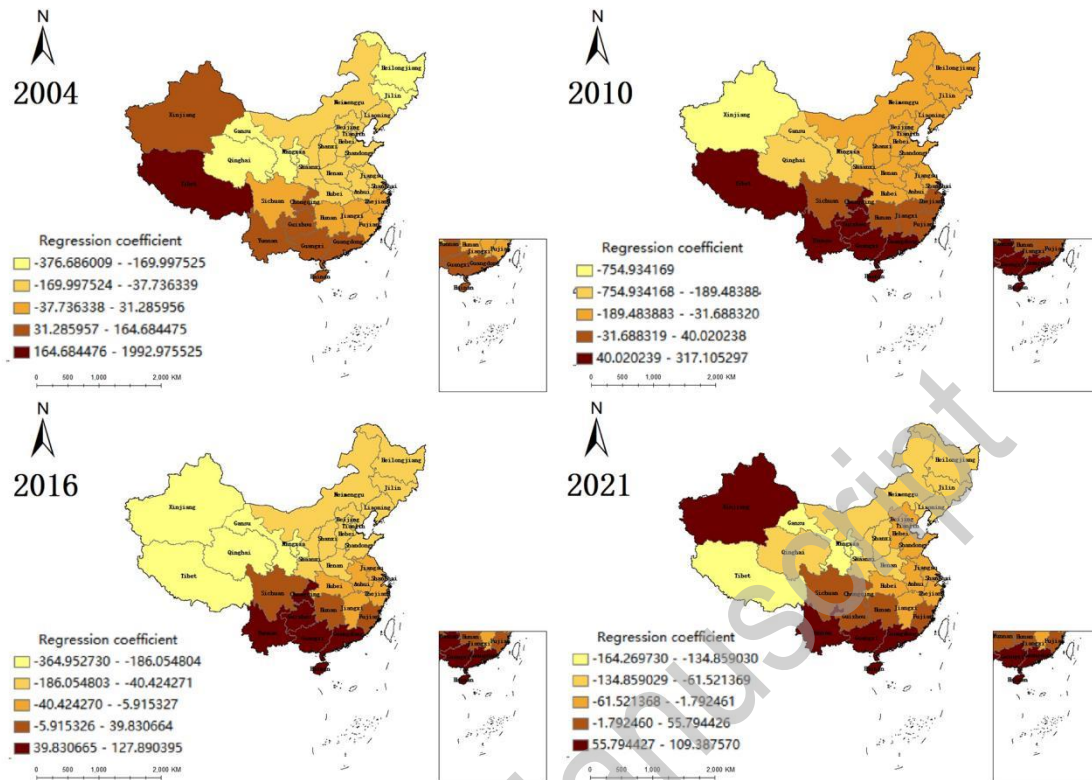
The estimated results of the GTWR model were obtained in Table5(*Supplementary material 2*).

Medicine and health, transportation, environment, economy, disease and educational quality of the regression coefficient of the average were -30.99, 38.77, -6.35, -65.54, 54.19 and -37.73, respectively. The influence of various macro factors on PTB was economy,

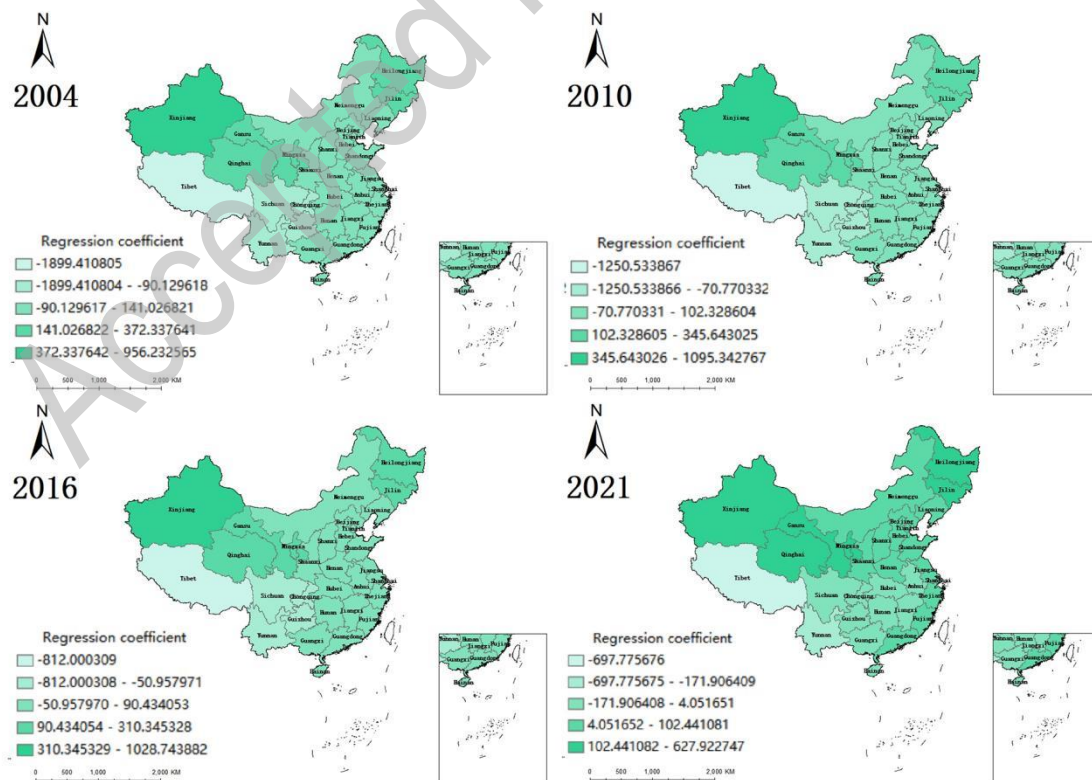
disease, transportation, educational quality, medicine and health, environment. The macro factors regression coefficient after visual situation was shown in the Figure.1 below.

Table5. Estimates of the GTWR model

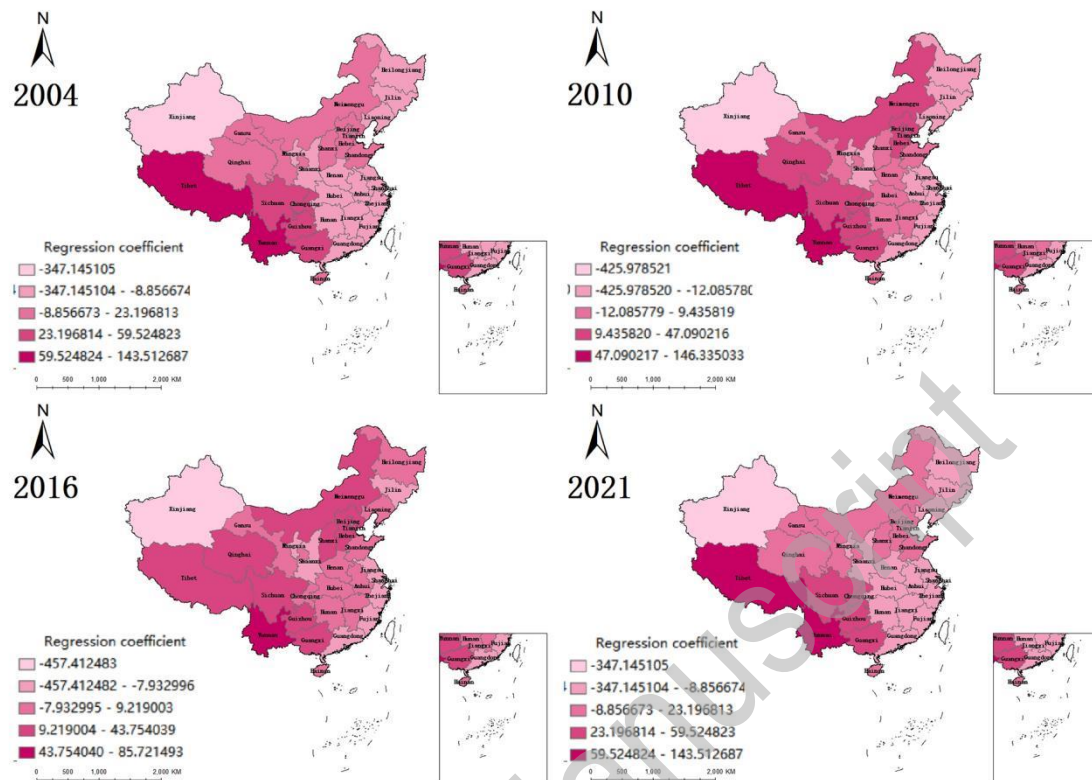
	Mean	Min	Lower quartile	Median	Upper quartile	Max
Intercept	101.95	41	80.8	97.72	123.5	285
Medicine and health	-30.99	-1021	-63.37	-38.51	30.13	1993
Transportation	38.77	-1899	-1.37	25.17	55.75	1106
environment	-6.35	-457	-12.67	-1.24	15.6	222
Economy	-65.54	-248	-110.72	-79.1	-48.75	248
Disease	54.19	-322	-25.71	62.82	146.6	272
Educational quality	-37.73	-863	-119.67	-45.4	10.03	1377



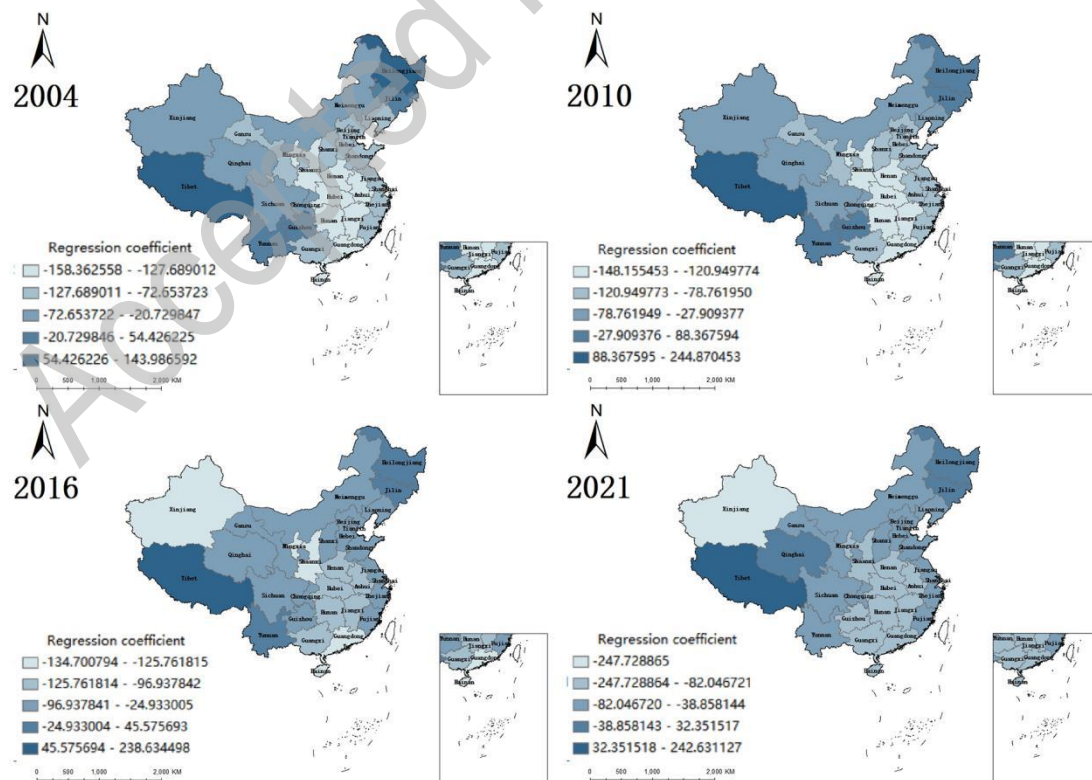
a the statistic distribution of medicine and health



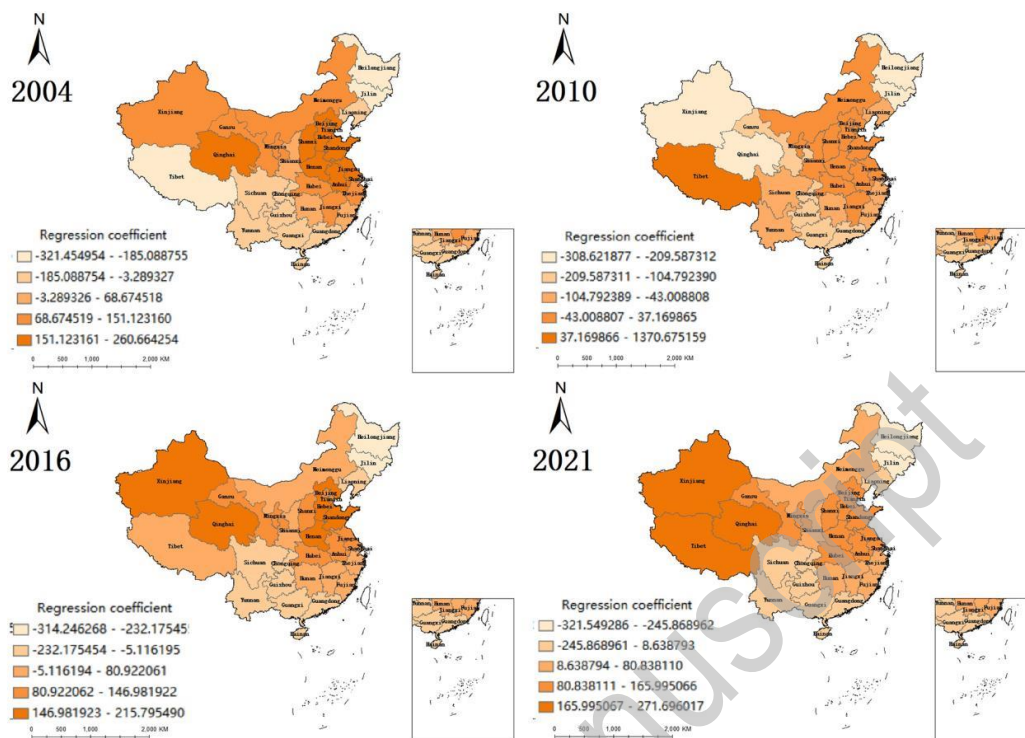
b the statistic distribution of transportation



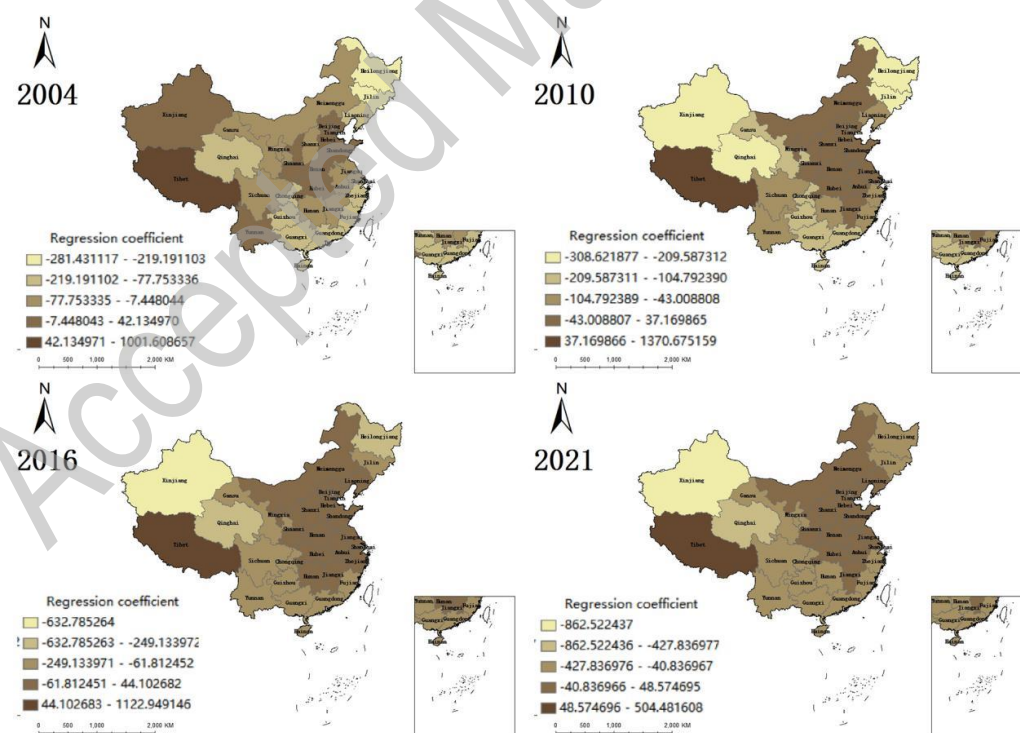
c the statistic distribution of environment



d the statistic distribution of economy



e the statistic distribution of disease



f the statistic distribution of educational quality

Figure1 2004, 2010, 2016, 2021 GTWR regression coefficient distribution

Discussion

To this day, China still has one of the highest PTB burdens. The analysis on the influencing factors of PTB was mostly focused on a single province or city and other local areas. This paper conducted modeling analysis on the incidence and influencing factors of PTB at the provincial level throughout the country from 2004 to 2021, and found that compared with the traditional OLS model and GWR model, the overall fitting effect of the GTWR model was better.

The results of GTWR model showed that different regions were affected by different macro factors to different degrees, that is, the incidence of PTB in China was spatially heterogeneous in different provinces and cities. From 2004 to 2021, the principal factors affecting the incidence of pulmonary tuberculosis in China are economy and disease. In most areas of China, economy was negatively correlated with the incidence of PTB, while disease was positively correlated with the incidence of PTB. Among them, Xinjiang Uygur Autonomous Region[11, 12, 13, 14] was highly affected by economic and disease factors, and the influence degree was increasing. This may be related to the shortage of local medical expenditure in areas with relatively backward economic development[15, 16, 17, 18], resulting in delayed diagnosis and treatment. The secondary factors affecting the incidence of PTB in China were the transportation and the educational quality. The transportation factors had a significant difference in the incidence of PTB in different areas, while the educational quality factors had a negative correlation with the incidence of PTB. We found that in Northwest, Northeast and North China, traffic indicators were positively correlated with the incidence of PTB, possibly because population flow increased the risk of spreading PTB bacilli, while in other regions with negative correlation, the most significant impact was in Southwest China, which may be related to the development of

local transportation driving the development of tourism and thus promoting the development of economy. The awareness of health screening and regular physical examination may be poor in areas with low education level, and early detection, diagnosis and treatment of diseases cannot be achieved. Educational status and health awareness among tuberculosis patients can influence their lifestyles in order to improve their living environments to prevent the spread of the infectious disease[19]. Health education interventions and efforts are needed to strengthen precise information dissemination to promote knowledge, attitude, and practices regarding tuberculosis among patients and non-patients at primary health care facility. In addition, the medical and health care and the environment also had a certain impact on the national incidence of PTB, that is, the lower the medical and health care and the more serious the environmental pollution, the higher the incidence of PTB[20, 21,22]. This suggest that we should strengthen the medical and health infrastructure, improve the medical service system, increase the number of designated medical institutions for PTB, and do a good job in environmental protection.

The incidence of PTB in China had spatial-temporal heterogeneity and was affected by the economy, disease, transportation, educational quality, medical and health care, environment, etc. It was characterized by a large difference between the north and the south in space and between the east and the west. In terms of time, the influencing degree of each factor was also different. In addition, this study also had some shortcomings, such as taking provincial administrative regions as the unit and failing to be accurate to prefecture-level cities. However, this study had a longtime span and a wide range of influencing factors, which can accurately reflect the temporal and spatial trend of PTB epidemic.

Therefore, the following suggestions are put forward: first, Qinghai, Tibet, Gansu and other

northwest regions should speed up economic construction, strengthen medical and health infrastructure construction, improve the medical service system, improve the diagnosis rate of PTB, and reduce under-reporting. Second, strengthen the construction of transportation in rural areas, expand the construction of roads and railways, and make it easier for residents in rural and remote areas to see a doctor in a timely manner. At the same time, for areas with developed traffic, check the floating population in time to prevent the flow of people from being too dense. Third, in southwest and North China, China will expand publicity, raise awareness of household waste classification and treatment, improve urban greening, reduce the emission of harmful substances in waste gas, and improve air quality. In areas with high incidence of PTB, health publicity and education on infectious diseases should be carried out in communities and schools, at the same time, make the integration of knowledge, belief and practice. In particular, ethnic minority gathering area should strengthen publicity, improve the awareness of PTB.

Conclusions

The influencing factors of GTWR model on the incidence of PTB in China are comparatively reasonable, which reflects that the incidence of PTB has spatial-temporal heterogeneity. The incidence of PTB is mainly affected by six macro factors, namely medicine and health, transportation, environment, economy, disease and educational quality.

Data availability statement

The extracted data can be found in the supplementary materials.

Acknowledgments

The authors thank National Bureau of Statistics in China for providing free and relevant datas about this study for us.

Author contribution

Conceptualization: L.S. and H.Y.; methodology: L.S. and H.Y.; Data collection: H.Y.; formal analysis: H.Y.; figure and table: H.Y., J.Y., Y.Y., H.Z. and Q.C.; Manuscript: H.Y., J.Y., Y.Y., H.Z., and Q.C.; Supervision: L.S.; All authors read and approved the final manuscript.

Financial support

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

Conflicts of Interest None.

References

1. **World Health organization.**(2022) Global tuberculosis report 2022. <https://www.who.int/zh>.
2. **Asemahagn MA-O, Alene GD, Yimer SA.**(2021) Spatial-temporal clustering of notified pulmonary tuberculosis and its predictors in East Gojjam Zone, Northwest Ethiopia. *PloS one*. 2021;16(1):e0245378.
3. **Ladha N,et al.**(2022) Determinants, risk factors and spatial analysis of multi-drug resistant pulmonary tuberculosis in Jodhpur, India. *Monaldi archives for chest disease*. 2022;92:2026.
4. **Guo Z,et al.**(2019) Epidemiological characteristics of pulmonary tuberculosis in mainland China from 2004 to 2015: a model-based analysis. *BMC Public Health*. 2019;19(1):219-229.
5. **Yang Y, et al.**(2016) Seasonality Impact on the Transmission Dynamics of Tuberculosis. *Computational and mathematical methods in medicine*. 2016;8713924. doi:10.1155/2016/8713924
6. **Guo B,et al.** (2021) Determining the effects of socioeconomic and environmental determinants on chronic obstructive pulmonary disease (COPD) mortality using geographically and temporally weighted regression model across Xi'an during 2014-2016. *The Science of the total environment*. 756, 143869-143883.
7. **Wang ZA-O,et al.**(2022) Identifying the Determinants of Distribution of *Oncomelania hupensis* Based on Geographically and Temporally Weighted Regression Model along the Yangtze River in China. *Pathogens (Basel, Switzerland)*. 11: 970-987.
8. **Yi S,et al.**(2021) Spatial and Temporal Characteristics of Hand-Foot-and-Mouth Disease and Its Response to Climate Factors in the Ili River Valley Region of China. *International journal of environmental research and public health*. 18, 1954-1962.
9. **Ge L, et al.**(2016) Construction of a Seasonal Difference-Geographically and Temporally Weighted Regression (SD-GTWR) Model and Comparative Analysis with GWR-Based Models for Hemorrhagic Fever with Renal Syndrome (HFRS) in Hubei Province (China). *International journal of environmental research and public health*. 13: 1062-1075.
10. **Huang B, Wu B, Barry M.**(2010) Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*. 2010;24:383-401.
11. **González-Domenech CM, et al.**(2021) A Pandemic within Other Pandemics. When a Multiple Infection of a Host Occurs: SARS-CoV-2, HIV and Mycobacterium tuberculosis. *Viruses*. 13: 931-943.
12. **Kerkhoff AD, Havlir DV.**(2021) Tuberculosis, Opportunistic Infections, and COVID-19 Among People with HIV. *Topics in antiviral medicine*. 29:344-351.
13. **MacNeil A,et al.**(2020) Global Epidemiology of Tuberculosis and Progress Toward Meeting Global Targets - Worldwide, 2018. *MMWR. Morbidity and mortality weekly report*. 69(11), 281-285.
14. **Riou C,et al.** (2021) Relationship of SARS-CoV-2-specific CD4 response to COVID-19 severity and impact of HIV-1 and tuberculosis coinfection. *The Journal of clinical investigation*. 131(12), e149125- e149139.
15. **He WC,et al.**(2020) Spatial inequality, characteristics of internal migration, and pulmonary tuberculosis in China, 2011-2017: a spatial analysis. *Infectious diseases of poverty*. 9(1), 159-

172.

16. **Liu MY,et al.** (2018)Spatial and temporal clustering analysis of tuberculosis in the mainland of China at the prefecture level, 2005-2015. *Infectious diseases of poverty*.7(1), 106-115.
17. **Mijiti P,et al.**(2016) Prevalence of pulmonary tuberculosis in western China in 2010-11: a population-based, cross-sectional survey. *The Lancet. Global health*. 4(7), e485–e494.
18. **Zhang Y,et al.** (2019)Spatial distribution of tuberculosis and its association with meteorological factors in mainland China. *BMC infectious diseases*.19(1), 379-345.
- 19.**Miandad,et al.** (2016). Educational status and awareness among tuberculosis patients of Karachi. *JPMA. The Journal of the Pakistan Medical Association*. 66(3), 265–269.
20. **Wang R,et al.**(2022)Spatial Heterogeneity and Its Influencing Factors of Syphilis in Ningxia, Northwest China, from 2004 to 2017: A Spatial Analysis. *International journal of environmental research and public health*.19(17),10541-10552.
21. **Xia L,et al.**(2020) Spatio-temporal analysis of socio-economic characteristics for pulmonary tuberculosis in Sichuan province of China, 2006-2015.*BMC infectious diseases*. 20(1),433-444.
22. **Yang J,et al.** (2020)A study on the relationship between air pollution and pulmonary tuberculosis based on the general additive model in Wulumuqi, China. *International journal of infectious diseases*. 96:42-47.