Health & Place 90 (2024) 103353

Contents lists available at ScienceDirect

Health and Place

journal homepage: www.elsevier.com/locate/healthplace

Uncovering spatiotemporal development patterns of AIDS in China: A study using panel data with Joinpoint Regression analysis and Spatial Clustering

Shu-nan Gui [a,1], Xiang Zhang [b,1], Zhenhui Sun [c,d], Yao Yao [b,e,*]

[a] Xi'an Jiaotong University Health Science Center, Xi'an Jiaotong University, 10049, PR China
[b] UrbanComp Lab, School of Geography and Information Engineering, China University of Geosciences, Wuhan, 430078, Hubei province, PR China
[c] Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University, 500 Dongchuan Road, Shanghai, 200241, PR China
[d] Key Laboratory of Spatial-temporal Big Data Analysis and Application of Natural Resources in Megacities (Ministry of Natural Resources), School of Geographic Sciences, East China Normal University, 500 Dongchuan Road, Shanghai, 200241, PR China
[e] National Engineering Research Center of Geographic Information System, China University of Geosciences, Wuhan, 430078, Hubei province, PR China

ARTICLE INFO

Keywords:
JPR analysis
AIDS prevalence rate
Clustering
Spatiotemporal feature extraction
Hotspot analysis

ABSTRACT

The HIV/AIDS epidemic in China is severe and complex. Comprehensive spatiotemporal analysis provides valuable insights for intervention policy formulation. Previous studies often overlooked local changes in time trends and regional disease development patterns. In this study, we propose a new spatiotemporal analysis method based on the Joinpoint Regression (JPR) model and K-means clustering to refine the division of stages in China's AIDS epidemic and differentiate geographical areas based on development patterns. We then use hotspot analysis to describe the current status of AIDS, presenting a comprehensive view of the epidemic in China from 2004 to 2018. JPR results show China's AIDS incidence generally increased during 2004–2018 (AAPC = 23.2), with a significant turning point in 2012. Time series feature clustering classifies the country into three regions: Southwest, Central and Eastern, and the other region. Each region corresponds to different epidemic causes and transmission pathways, informing targeted interventions. Hotspot analysis reveals the Southwest region as the most severely affected area, requiring intensified prevention and control efforts. This study offers a novel from both time and space dimensions for understanding and combating the AIDS epidemic, furnishing valuable references for policymakers in the further development of strategies.

1. Introduction

HIV/AIDS, a chronic infectious disease that seriously affects global public health (Qiao et al., 2019), is the direct cause of nearly one million deaths each year (collaborators, 2019; Xu and Zhan, 2024). The Joint United Nations Programme on HIV/AIDS (UNAIDS) has established a series of objectives aiming to end the HIV/AIDS epidemic by 2030 (Ghys et al., 2018). For developing countries, HIV/AIDS has always been a significant challenge for local public health (Bonovas and Nikolopoulos, 2012; Campbell et al., 2017). Since the first case of HIV/AIDS was reported in China in 1985, the number of HIV/AIDS-related deaths has been significantly higher than that of other infectious diseases (Li et al., 2014), and its morbidity and transmission in different regions show significant differences (Huang et al., 2022). The development pattern of HIV/AIDS in China is complex and exhibits similarities to both developed and developing countries (Basavaraju et al., 2010; Gokhale et al., 2017; Xu et al., 2021), which can provide insight into the global AIDS epidemic.

Similar to the epidemiological research of other diseases, analyzing the spatiotemporal characteristics of HIV/AIDS can help obtain essential information about its epidemiology, monitoring, and detection (Goovaerts, 2013; Wang et al., 2021), which is of great significance for understanding and controlling the disease. Based on temporal patterns, the HIV/AIDS epidemic in China can be divided into three stages: sporadic epidemic stage (1985–1988), local cluster epidemic stage (1989–1994), and widespread epidemic stage (1995 to present) (Wang,

---

* Corresponding author. UrbanComp Lab, School of Geography and Information Engineering, China University of Geosciences, Wuhan, 430078, Hubei province, PR China.

E-mail addresses: 359346472gsn@stu.xjtu.edu.cn (S.-n. Gui), z.xiang@cug.edu.cn (X. Zhang), sunzhenhui@stu.ecnu.edu.cn (Z. Sun), yaoy@cug.edu.cn, yao.yao@urbancomp.net (Y. Yao).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.healthplace.2024.103353
Received 18 June 2024; Received in revised form 2 September 2024; Accepted 9 September 2024

2007; Xu et al., 2021), showing apparent changes in development trends. Along with the changing trends in the HIV/AIDS incidence rate, its epidemic characteristics have also become more complex (Qian et al., 2014). For example, the number of HIV-infected cases in Yunnan Province has begun to decrease in recent years, while the incidence in Sichuan Province has risen sharply between 2015 and 2017, and the transmission mode has shifted from blood transmission associated with drug trafficking to predominantly sexual transmission (Jia et al., 2010). This change in epidemic characteristics is closely related to the spatial distribution of HIV/AIDS within China and may be influenced by various factors such as rural population migration to cities and gender ratio imbalance (He, 2007; Huang et al., 2022). By analyzing time and space, we can obtain a more comprehensive understanding of the spatiotemporal distribution characteristics and development patterns of HIV/AIDS.

To investigate the development trends of AIDS, it is essential first to analyze the changes in the incidence rate over time. Currently, widely used time series research models include regression analysis, autoregressive integrated moving average (ARIMA) model, and segmented regression model (Gray et al., 2021; Hincapie-Castillo and Goodin, 2023; Li et al., 2021; Niu et al., 2020). Traditional time series analysis studies indicate that conventional regression techniques can only reflect the overall trends of the data and not the local trend changes, resulting in limitations in fully capturing the changing trends of time series (Gray et al., 2021). The ARIMA model is a common method for analyzing time trends, capable of identifying and controlling non-stationary patterns in the data (Niu et al., 2020). However, this method necessitates a considerably larger quantity of data, at least 50 pre-intervention observations for a standard interventional ARIMA (Box and Tiao, 1975). Furthermore, the method requires pre-determined intervention nodes and is only capable of assessing the impact of a single intervention (Li et al., 2021). In addition to the methods mentioned above, some researchers have employed segmented regression techniques (Hincapie-Castillo and Goodin, 2023; Kim et al., 2000).

Segmented regression techniques can estimate instantaneous changes in trends over time and reveal specific trends in local data. However, as this technique necessitates the determination of segmentation points based on the assumption of trend turning points, it cannot explain some non-stationary patterns in the data, such as seasonality and cyclicity (Li et al., 2021). JPR is extensively applied in the fields of cancer and epidemiology for analyzing time trends and characteristics, as well as evaluating the impact of intervention measures on the incidence of AIDS (Ma et al., 2021; Hincapie-Castillo and Goodin, 2023; Kim et al., 2000).

To further investigate the prevalence characteristics of AIDS, spatial epidemiology, which combines Geographic Information Science (GIS) and spatial analysis techniques, is increasingly being applied in epidemiological research (Carrel et al., 2016). In 2008, Ezekiel Kalipeni (Kalipeni and Zulu, 2008) utilized spatial interpolation methods within GIS to address the lack of AIDS data in African regions, demonstrating the utility of GIS methods in spatiotemporal analysis for HIV prevalence research. Additionally, health geographic information systems play a crucial role in the research of disease clustering (Kandwal et al., 2009), as the identification and classification of spatial aggregations of disease incidence facilitate public health monitoring efforts.

A key research area in spatial epidemiology is the spatial distribution characteristics of diseases. Wang investigated the impact of varying levels of urbanization in different regions on the incidence of obesity among residents, demonstrating a certain correlation between regional environment, lifestyle, and disease incidence (Wang et al., 2020a). Additionally, Relevant scholars identified the main factors and their changes influencing the spatial distribution of foodborne diseases using the Spatial Lag Panel Regression Model (SLM) and the Geographically and Temporally Weighted Regression Model (GTWR) (Wang et al., 2020b). Leo employed GIS interpolation methods to extract regional-level prevalence estimates and analyzed the spatial dependence of HIV prevalence (Zulu et al., 2014). Earlier studies have also investigated the spatial correlation of the AIDS epidemic on global and local scales using Moran's I index (Huang et al., 2022). However, these studies focused on spatial analysis at specific time points without incorporating time series characteristics, while time series analysis can help uncover potential patterns in epidemic changes (Hamilton, 2020).

Consequently, the analysis of spatiotemporal characteristics is crucial. Space-time scan statistic represent a classic method in this field (Kulldorff et al., 1998), often used in epidemiology to detect spatiotemporal clusters of anomalous results in both temporal and spatial dimensions, in other words, newly emerging disease clusters. Employing Joinpoint Regression (JPR) allows for the examination and comparison of time trend changes within different clusters. By combining JPR analysis with K-means clustering and hotspot analysis, further investigation into the geographical differences in the development patterns and incidence status of AIDS in China is conducted. This approach allows for a more accurate and comprehensive assessment of the changes in the epidemic.

This study collected AIDS incidence data for China and its 31 provinces from 2004 to 2018, using a publicly available panel dataset, aiming to provide a detailed exploration of the time trend of AIDS in China and fill the gap in the research field of AIDS development patterns. A novel method was proposed. Initially, the JPR model was employed for stage-wise fitting of AIDS incidence, identifying turning points in disease progression and estimating the immediate changes in trends over time. Subsequently, this study utilized the time series in the original AIDS incidence curve to calculate the numerical characteristics of the sequence data and the curve trend shape features. K-means clustering method was applied to spatially cluster these features, identifying spatial differences in the development patterns of AIDS in China. This study then discusses the practical implications of spatiotemporal development patterns. Furthermore, a hotspot analysis of the latest AIDS incidence was conducted, examining the current prevalence of AIDS in China and offering suggestions for future prevention and control efforts.

2. Data sources

The area of this study is China, which is the world's largest developing country and has the largest population. Since the reform and opening up, China has experienced significant transformations in various aspects, such as its economy, society, and culture. Studies on the spread and development of AIDS in China can offer valuable insights for other developing countries worldwide.

The AIDS incidence data used in this study is sourced from the publicly available compulsory reporting monitoring data provided by the China Public Health Science Data Center (www.phsciencedata.cn). This data center is an external data service portal established by the Chinese Center for Disease Control and Prevention. The database has recently undergone updates under strict quality assessment and has been utilized in multiple studies (Tao et al., 2020; Lin et al., 2023), making it a reliable data source. As of the time of writing, the database contains AIDS incidence data from 2004 to 2018. Consequently, this study collected AIDS incidence data for China and its 31 provinces from 2004 to 2018 (Table 1).

Table 1
Partial display of raw AIDS prevalence (1/100,000) data.

Year	China	Beijing	…	Ningxia	Gansu
2004	0.235	0.479	…	0.097	0.130
2005	0.432	0.495	…	0.068	0.105
…	…	…	…	…	…
2017	4.145	3.433		1.497	1.843
2018	4.620	3.907	…	1.687	2.354

3. Methodology

The methodology of this study comprises three parts: (1) JPR analysis. This section employs JPR to fit the AIDS incidence data in China, obtaining the fitting curves and crucial indicators reflecting fitting accuracy and temporal trends, such as MSE, APC, and AAPC, for China and its provinces. (2) AIDS development pattern cluster analysis. This section focuses on the time series fitting curves of incidence rates for each province, calculating 30 curve feature values, including mean, variance, skewness, and kurtosis, to quantify the characteristics of the time series data. The curve feature value results are then subjected to K-means clustering, classifying the AIDS development patterns spatially. (3) Hotspot analysis. This section conducts hotspot analysis using the 2018 AIDS incidence data, determining the cold and hot spot spatial distribution of AIDS in China in 2018 to reflect the current state of the epidemic. The technical roadmap of this study is illustrated below (Fig. 1).

3.1. *JPR-based time series analysis*

The fundamental principle of fitting the JPR model involves dividing a long-term trend into several statistically significant segments and characterizing them. Depending on the data type, the model can be categorized into linear data models ($y = xb$) and log-linear data models ($\ln y = xb$) (Kim et al., 2000). A linear model is selected when the dependent variable follows a normal distribution (or an approximate normal distribution) and the sample size is large (typically greater than 100). In contrast, a log-linear model is suitable when the dependent variable follows an exponential or Poisson distribution, especially when the disease is rare or the rates come from a small population.

3.1.1. *Grid Search Method (GSM) modeling*

The JPR method used in this study is based on the Joinpoint Trend Analysis Software implementation (https://surveillance.cancer.gov/joinpoint/) (Kim et al., 2000), which employs the Grid Search Method (GSM) for modeling. The principle is to determine all possible interval piecewise function Joinpoints through the GSM and calculate the Sum of Squares Errors (SSE) and Mean Squared Errors (MSE) for each possible scenario. Then, the grid point with the smallest MSE is selected as the Joinpoint of the piecewise function, and the equation parameters, such as $\beta_0, \beta_1, \delta_0, \ldots \delta_k$ are fitted according to the selected Joinpoints and interval functions.

3.1.2. *Monte Carlo permutation test model selection*

The Monte Carlo permutation test is the default model selection method in Joinpoint Trend Analysis Software. In addition to the Monte Carlo permutation test, the software also provides the Bayesian Information Criterion (BIC), Modified BIC (MBIC), and related derived methods for model selection, but the latter is mainly for advanced users with special needs and is not suitable for this study. The applications have shown that the permutation test approach worked well for cancer incidence and mortality data. This study chooses the default Monte Carlo permutation test.

3.1.3. *Indicators of accuracy and trends*

The software reflects the accuracy through the *t*-test of the fitting curve parameters and the accuracy index Mean Squared Error (MSE) of the fitting curve:

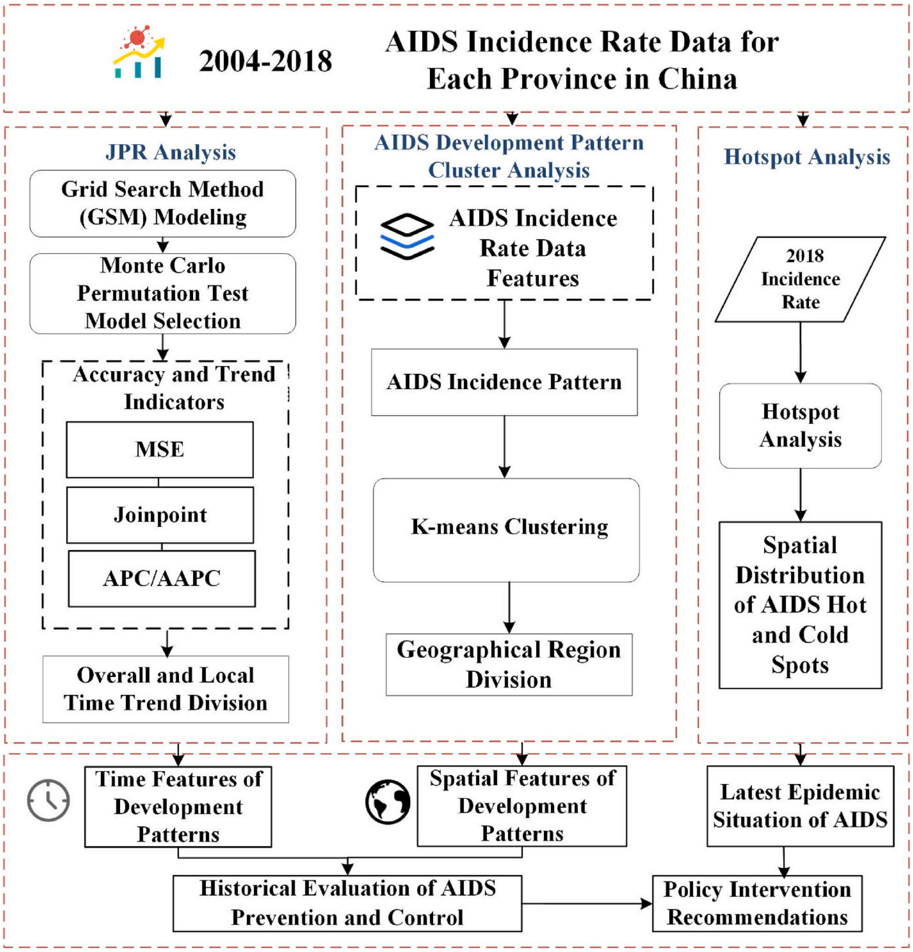


Fig. 1. The technical roadmap of the study, including three parts: (1) JPR Analysis; (2). AIDS development pattern cluster analysis; (3) Hotspot analysis.

$$MSE = \frac{1}{n}\sum\nolimits_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \qquad \text{eq.1}$$

Joinpoints and their number are also intuitive indicators in the results, which determine the division of the result curve stages, and their different numbers reflect the heterogeneity of different fitting curves. The annual percentage change (APC), average annual percentage change (AAPC), and their 95% confidence interval (CI) are the main result indicators of JPR model. APC is the percentage of the average annual change in the dependent variable, used to evaluate the internal trend of each independent interval of the piecewise function or the global trend when the number of Joinpoints is 0. AAPC is used to evaluate the global average change trend containing multiple intervals comprehensively. The APC and AAPC can be used to evaluate the trend of AIDS incidence in China. If the APC is greater than 0, it indicates that the incidence is increasing year by year; otherwise, it is decreasing year by year. If there are no Joinpoints, then APC = AAPC, indicating that the data trend shows a unidirectional increase or decrease trend (Wei et al., 2017).

3.2. Cluster analysis of AIDS time series based on K-means algorithm

The original time series data on AIDS incidence reveals the trends of changes in AIDS incidence rates in different regions over time. These trends, to some extent, reflect the transmission patterns of AIDS in these areas. Analyzing the time series of AIDS incidence rates plays a crucial role in studying its patterns of occurrence.

In the analysis of time series, a common method is to calculate its curve characteristics, thereby quantifying the features of the time series data. This technique facilitates a comprehensive analysis of the time series as a whole rather than individual data points (Wang et al., 2006). In addition, for time series data, entropy and complexity can be used to describe and identify the inherent regularities of the dynamical system (Hansen, 1995). In this study, this corresponds to the development patterns of AIDS in different provinces. For the time series of AIDS incidence rates in each province from 2004 to 2018, 30 curve characteristics, including mean, variance, skewness, kurtosis, etc., are calculated to quantitatively describe the characteristics of AIDS incidence in the entire region during this period. Some of the curve characteristics are listed in Table 2, and details can be found in Appendix Table S1.

Table 2

The time series curve characteristics calculated in this study.

Feature Type	Feature Name	Mathematical Formula
Numerical Feature	Mean	$Mean = \frac{1}{n}\sum\nolimits_{i=1}^{n} f(x_i)$
	Median	$f(x)_{mid}$
	Standard Deviation	$\sigma = \sqrt{\frac{\sum\nolimits_{i=1}^{n}\left(f(x_i) - Mean\right)^2}{n-1}}$
	Range	$Range = f(x)_{max} - f(x)_{min}$
	…	…
Trend Feature	Distance to the i-th peak	$Distance_i = peak_i - x_0$
	Area to the i-th peak	$S_i = \int_0^{peak_i} f(x)dx$
	Length to the i-th peak	$L_i = \int_0^{peak_i}\sqrt{1 + f(x)'^2}\,dx$
	Skewness	$Skewness = \frac{n}{(n-1)(n-2)} \bullet \frac{\sum\nolimits_{i=1}^{n}(x_i - Mean)^3}{\sigma^3}$
	Burstiness statistic	$B = \frac{\sigma - m}{\sigma + m}$
	Kurtosis	$Kurtosis = \frac{n \cdot \sum\nolimits_{i=1}^{n}(f(x_i) - Mean)^4}{\left(\sum\nolimits_{i=1}^{n}(f(x_i) - Mean)^2\right)^2} - 3$
	Coefficient of Variation	$CV = \frac{\sigma}{Mean} \times 100\%$
	Shannon Entropy	$H(x) = \sum\nolimits_{i=1}^{n} p(f(x_i))\log_2 p(f(x_i))$
	…	…

Where x_i is the time corresponding to the i-th data point in the time series, $f(x_i)$ is the value at that time, n is the time series sample size, $peak_i$ is the time at which the i-th peak of the time series curve appears, and m represents the average time interval.

By analyzing the time series of AIDS incidence rates in different provinces, a quantitative description of the development of AIDS in each province can be made. To further analyze the national development pattern of AIDS, this study clusters the development patterns of AIDS in various regions, distinguishing the development patterns of AIDS in provinces nationwide into several categories. This distinction facilitates subsequent specific analysis of different categories of incidence patterns and the targeted formulation of preventive measures.

Clustering is one of the classic problems in the field of computer machine learning, and the K-means algorithm is widely used in clustering problems. It is a clustering algorithm based on Euclidean distance, assuming that the closer the distance between two objects, the greater the similarity. The main idea is: in a given dataset, first randomly select K initial cluster centers, use Euclidean distance to calculate the distances between each sample and divide the samples into K clusters based on the nearest distance principle; then calculate the average position of each cluster, reassign the cluster centers, and when the iteration reaches a specified number, or the cluster centers no longer change, the clustering ends (Wong, 1979).

This section first calculates the 30 curve characteristics of the AIDS incidence rate time series data for each province, which is used to describe the incidence pattern of AIDS quantitatively. Based on this, the K-means clustering method is used to cluster the AIDS development patterns of various provinces in China. This study explores the different AIDS development patterns existing in various provinces of China, analyzes the reasons for the different development patterns in different regions, and provides targeted prevention and control suggestions for different regions.

3.3. Latest AIDS situation analysis based on hotspot analysis

Hotspot analysis is a commonly used spatial analysis method in GIS, which can calculate the Getis-Ord Gi* statistic for each feature in a dataset. Through the obtained z-score and p-value, we can identify the locations where high-value or low-value features are spatially clustered (Getis and Ord, 1992). High-value features often attract attention, but they may not be hotspots with significant statistical significance. In order to be classified as a statistically significant hotspot, elements must exhibit high values and be situated in proximity to other elements that also possess high values. The calculation method of the Getis-Ord Gi* statistic is as follows:

$$G_i^* = \frac{\sum\limits_{j=1}^{n} w_{i,j}x_j - \overline{X}\sum\limits_{j=1}^{n} w_{i,j}}{S\sqrt{\left[n\sum\limits_{j=1}^{n} w_{i,j}^2 - \left(\sum\limits_{j=1}^{n} w_{i,j}^2\right)^2\right] \Big/ (n-1)}} \qquad \text{eq.2}$$

Where x_j is the AIDS incidence rate of the province, $w_{i,j}$ is the spatial weight between features i and j, and n is the total number of features.

$$\overline{X} = \frac{\sum\limits_{j=1}^{n} x_j}{n} \qquad \text{eq.3}$$

$$S = \sqrt{\frac{\sum\limits_{j=1}^{n} x_j^2}{n} - \left(\overline{X}\right)^2} \qquad \text{eq.4}$$

The Gi* statistic returned by each feature in the dataset is the z-score. For significant positive z-scores, the higher the z-score, the tighter the clustering of high values (hotspots). For statistically significant negative z-scores, the lower the z-score, the tighter the clustering of low values (cold spots). In this section, hotspot analysis is conducted on the AIDS incidence rates in various provinces of China in 2018 to evaluate and analyze the latest development trends of AIDS in China.

4. Results

4.1. Results of JPR-based fitting of AIDS incidence rates

4.1.1. JPR fitting accuracy for AIDS incidence rates

In this study, the JPR model is applied to fit the AIDS incidence rate in China between 2004 and 2018. The results, measured by the accuracy index MSE value, indicate that the MSE values for Xinjiang, Yunnan, and Guangxi provinces exceed 0.05, while those for the remaining provinces are below 0.05, demonstrating statistical significance (Appendix Table S2). These findings suggest that the JPR model exhibits strong applicability in fitting the trend of AIDS incidence rate and categorizing the stages of AIDS incidence.

4.1.2. AIDS incidence curve trends and segmentation

In this study, we applied the JPR model to fit the AIDS incidence rate in China between 2004 and 2018, determining the Joinpoint characteristics of the AIDS incidence rate curve (Fig. 2) and the trend change indicators APC and AAPC (Appendix Table S3). As shown in Fig. 2, results are classified into three categories based on the number of Joinpoints: Category I includes Shanxi, Henan, and Tibet provinces; Category II comprises Tianjin and 20 other provinces; Category III consists of Beijing and 8 other provinces.

The national AIDS incidence rate's AAPC is 23.2, indicating a growth trend from 2004 to 2018. As illustrated in Fig. 3, the AAPC values for all 31 provinces in China are above 0, suggesting an increasing trend in AIDS incidence rates across all provinces between 2004 and 2018. The southwestern regions, such as Sichuan, Chongqing, and Guizhou, experienced the most significant growth, while central regions, including Henan, Hubei, and Anhui, saw the most minor increases. Two factors may contribute to the continuous rise in China's AIDS incidence rate: the rampant sexual transmission led by MSM (Men who have sex with men) and the increased detection rate (Wu et al., 2019; Feng et al., 2009; Meng et al., 2018).

The APC of the national AIDS incidence rate was 34.71 (95% CI = 30.7 ~ 38.9) from 2004 to 2012 and 9.30 (95% CI = 7.2 ~ 11.57) from 2012 to 2018, suggesting a higher growth rate from 2004 to 2012 than

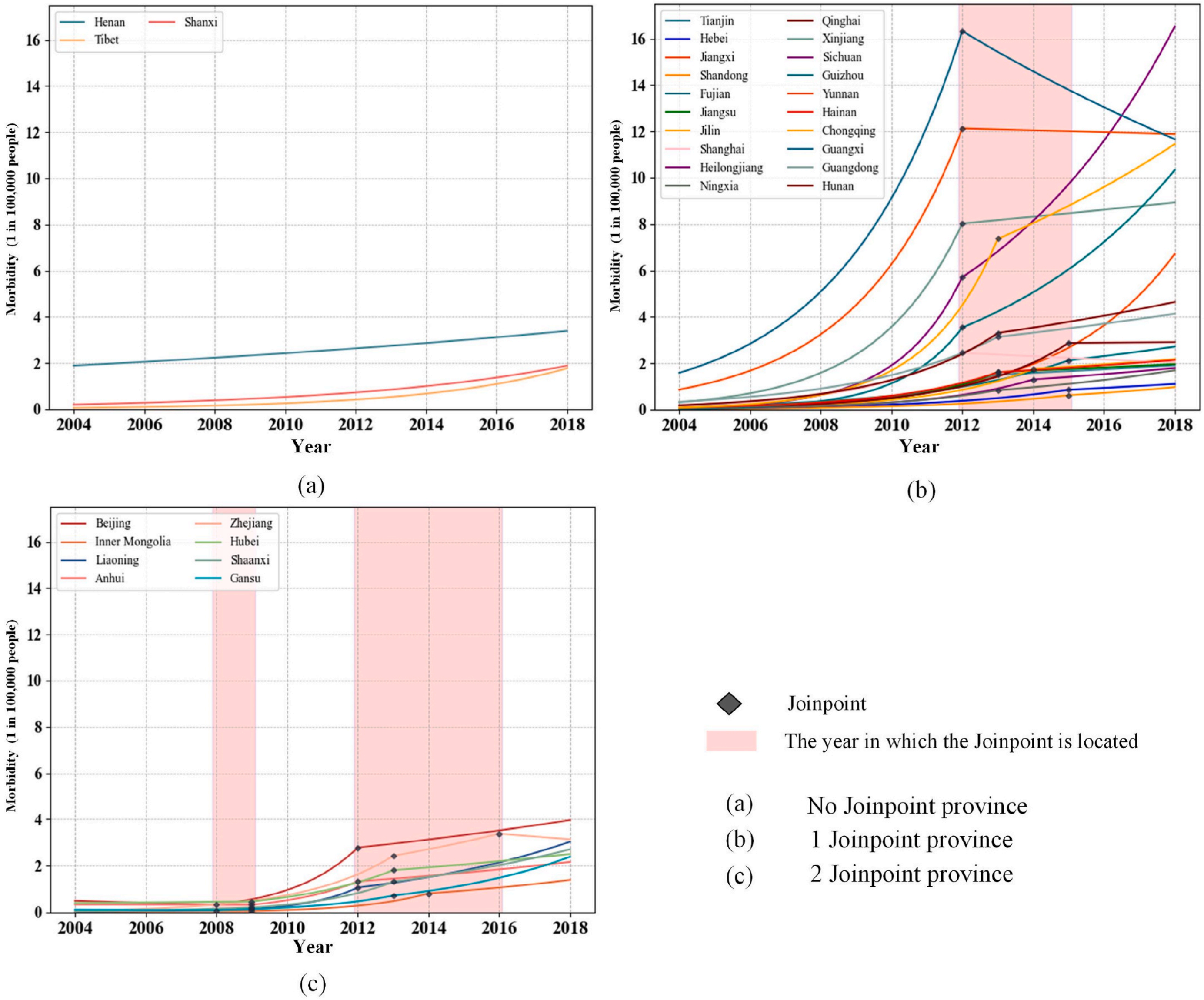


Fig. 2. JPR fitting results in three categories based on the number of Joinpoints.

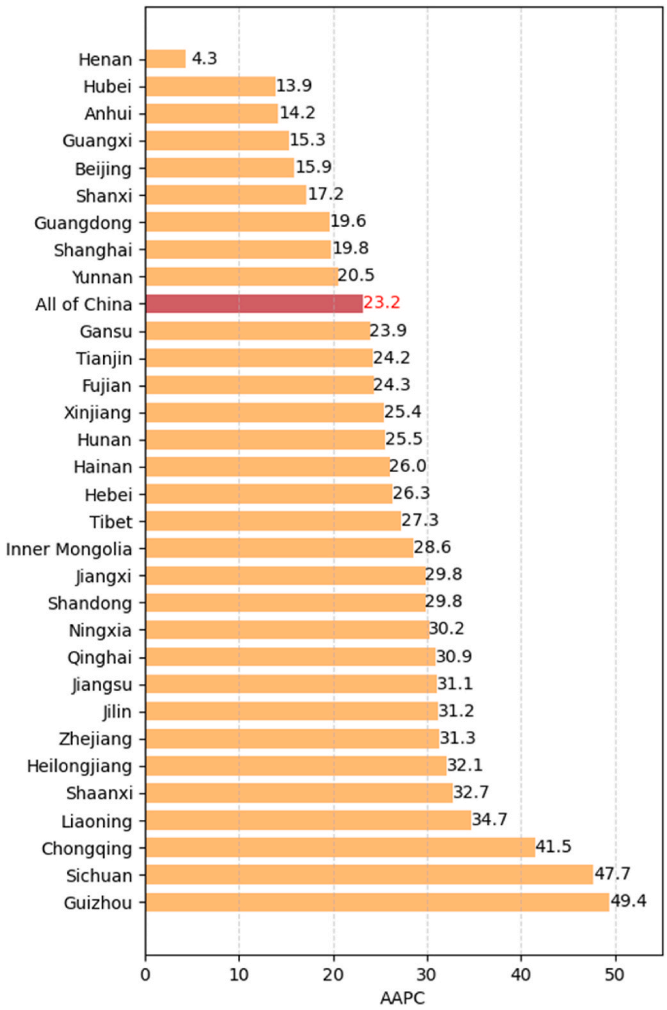


Fig. 3. Comparison of the AAPC results between various provinces and the entire country.

from 2012 to 2018, with the incidence rate growth slowing down after 2012. Appendix Table S3 displays the APC values and their confidence intervals for the fitted AIDS incidence rate curves of the 31 provinces and cities in China. In Categories II and III, Joinpoints are mainly observed between 2008-2009 and 2012–2015. The first Joinpoint in Category III occurs between 2008 and 2009, with the APC value increasing after the node, indicating an acceleration in incidence rate growth. The second Joinpoint in Category III and the Joinpoint in Category II are both concentrated between 2012 and 2015, with both showing a decrease in APC value after the Joinpoint, representing a slowdown in incidence rate growth. However, only Shanghai, Guangxi, Yunnan, and Zhejiang exhibit a declining trend in incidence rates after the Joinpoint, with Guangxi's result being significant (Appendix Table 2). Local policies formulated by the local government around 2013 (General office of the standing committee of the Guangxi Zhuang autonomous region people's congress, 2013), which are tailored to local conditions (Chen et al., 2018), are key to Guangxi's exceptional performance in epidemic prevention and control.

4.2. *Results of clustering of AIDS development patterns*

Following the application of the K-means clustering algorithm to the 30 time series curve features of each province, with the number of clusters set to 3 (with the exception of Taiwan Province, where no data was available), the resulting clustering outcomes were obtained (Fig. 5). Fig. 4 illustrates a geographical distribution pattern evident in the clustering results. Provinces belonging to the same category tend to be geospatially proximate, aligning with the first law of geography and the pattern of AIDS transmission. The clustering results demonstrate that provinces and cities with similar patterns of disease development are also spatially clustered. The 31 provinces and cities in China have been categorized into three types of AIDS development patterns: Southwest region, Central and Eastern region, and the other region. Different development patterns represent different transmission routes and times. Early drug transmission was dominant in the southwest region (Wu, 2015), while the Central and Eastern regions experienced an outbreak of blood-borne transmission in the last century (Dou et al., 2010); the AIDS epidemic in the other region mainly originated from the transmission in the first two categories of regions. Analyzing the development patterns of each category of provinces can provide valuable insights into the current prevention and control of AIDS in China.

4.3. *Results of AIDS hotspot analysis*

Based on the hotspot analysis of the incidence of AIDS in each province of China in 2018, the latest AIDS epidemic situation in China can be assessed. As shown in Fig. 6, the latest AIDS epidemic trend still exhibits significant regional differences and clustering. Seven hotspots are concentrated in the southwest of China, including Sichuan, Chongqing, Yunnan, Guangxi, Guizhou, Hunan, and Hainan, highlighting the severity and clustering of the AIDS epidemic in the southwest region. In contrast, ten cold spots are found in the eastern and northern regions, led by Beijing, Tianjin, Hebei, and Shandong, suggesting that the risk of AIDS in these areas is relatively low. This result indicates that the southwest region remains the area most severely affected by the AIDS epidemic in China, while the northeastern provinces have better controlled the epidemic.

5. Discussion

This study employs the JPR model and K-means clustering to propose a new spatiotemporal analysis method based on the China Public Health Science Database. By applying this method, this study analyzes the incidence of AIDS in China from 2004 to 2018 in both temporal and spatial dimensions. This method utilizes segmented regression technology, which effectively fits the time trend of the incidence rate and reveals immediate changes in the trend. The K-means clustering algorithm spatially divides the development patterns of AIDS in China. Building on the preceding work, a hotspot analysis of the AIDS incidence rate in 2018 can be used to determine the current situation of the AIDS epidemic in China.

The method proposed in this study has three significant advantages: (1) This method can identify turning points through JPR analysis and reveal trend changes between different stages, providing a powerful tool for evaluating intervention effects; (2) This method permits the differentiation of regional variations in disease development patterns and provides a comprehensive overview of the current state of the AIDS epidemic in China, which facilitates the formulation of targeted policies; (3) This method explores the spatiotemporal characteristics of AIDS in China more comprehensively in both temporal and spatial dimensions. By using this method, the spatiotemporal development patterns of AIDS in China from 2004 to 2018 can be more thoroughly investigated, deepening the understanding of the epidemic characteristics in different regions and providing support for precise interventions.

To address the issue of overlooking changes in trends within time series in previous studies, this study employs JPR model to unveil the temporal trend shifts in each stage, offering an objective basis for policy evaluation. Among all the MSE results, the values of Xinjiang, Yunnan and Guangxi exceeded 0.05. The MSE values were very sensitive to outliers (Chicco et al., 2021). These three provinces are located in the western part of China, where the AIDS epidemic is serious and the transmission patterns are complex with many outbreaks of the epidemic

Fig. 4. Clustering results of AIDS development patterns of Chinese provinces: Type1: Southwest region; Type2: Central and East region; Type3: The other region.

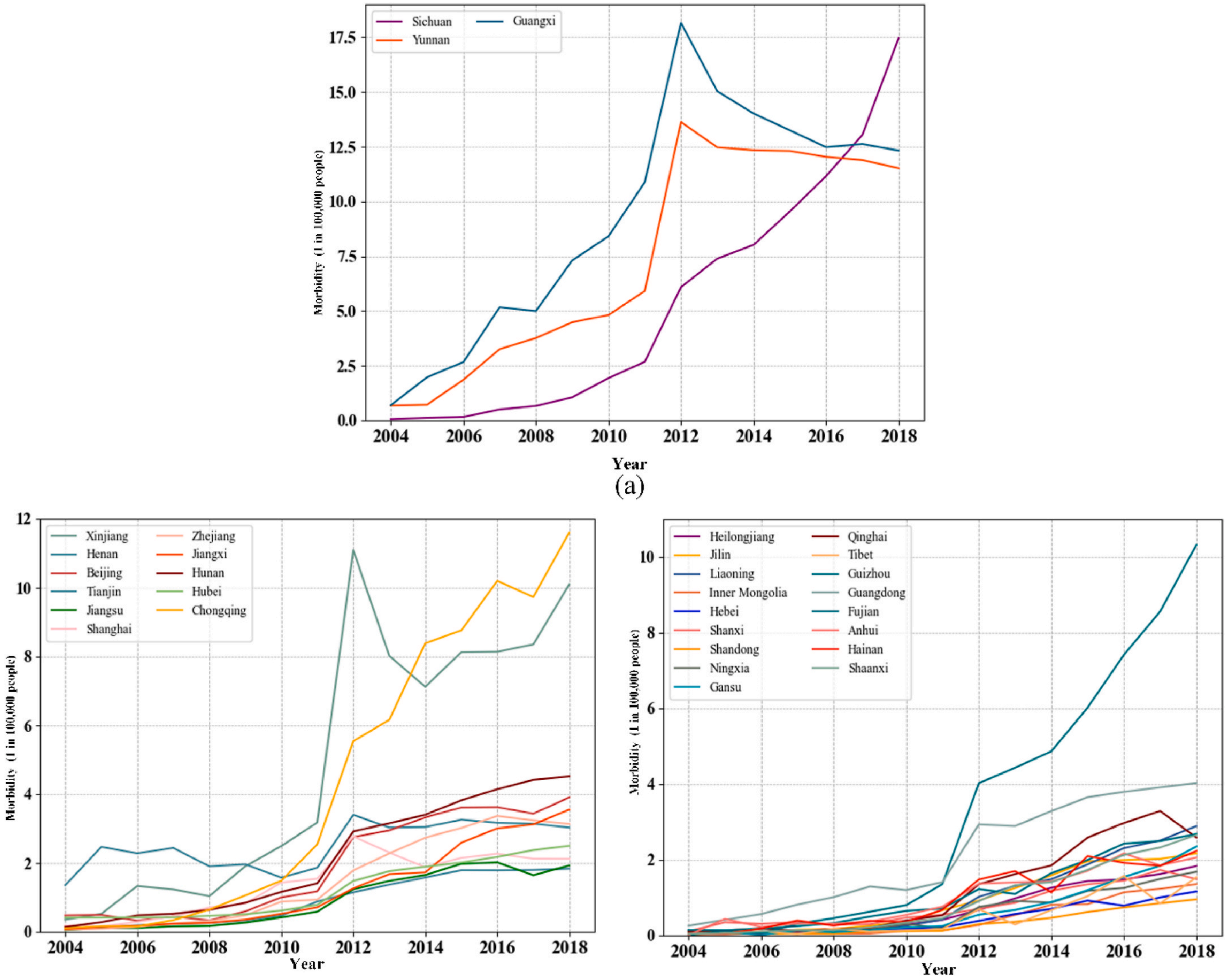
through drug and other transmission routes (Wu et al., 2019; Jia et al., 2010), which in turn affects the fitting accuracy. As indicated by the fitting outcomes of this study, AIDS incidence in China exhibited an overall ascending trend from 2004 to 2018 (Fig. 2), aligning with the conclusions of prior investigations (Xu et al., 2020; Wu, 2015). Following the Joinpoints between 2012 and 2015, the rate of AIDS progression across various regions collectively decelerated. This period corresponds with the execution of China's 12th Five-Year Plan (State Council Office Of The People's Republic Of China, 2012; PRC National Health, 2013) and the ongoing policy interventions (Lau et al., 2011; Zeng et al., 2016; Wang et al., 2018; Jiang et al., 2018). The decline in the incidence growth rate signifies the efficacy of these interventions.

The Joinpoint outcomes innovatively partition the AIDS incidence trend in China into distinct phases, objectively quantifying the instantaneous changes in the trend. By scrutinizing the alterations in the incidence growth rate, the effectiveness of disease interventions can be intuitively assessed, further showcasing the merits of the JPR model in policy impact evaluation. For instance, only Guangxi Province exhibited a declining trend in incidence rate post-2012, emphasizing the exceptional outcomes of localized policy interventions.

To address the deficiencies in the regional distribution of AIDS development patterns, this study applies the K-means algorithm for clustering, revealing the heterogeneity of AIDS development patterns in different regions of China and providing support for precise intervention. In the case of the results from Xinjiang province, previous clustering studies often grouped the province with provinces in the southwest from the perspective of high incidence rates (Qian et al., 2014; Huang et al., 2022). However, in the results of this study, the pattern of epidemic development is categorized as central and eastern regions. This phenomenon may be due to the fact that drugs in Xinjiang and the central and eastern region mainly originate from the inflow from places such as Sichuan (Xing et al., 2014), and the epidemic development pattern will also differ from provinces with direct drug input, such as the southwestern region. Compared with the direct application of spatial analysis of the incidence rate or number of cases, the analysis of the development pattern of the epidemic can effectively clarify the development pattern and causes of the disease and help to block and prevent the epidemic from the transmission path. It also provides reference and guidance for similar outbreaks that may occur in the future.

The results of this study suggest that policy intervention plays a significant and visible role in controlling the AIDS epidemic. Although no cure has been found to eradicate AIDS, policy support for the prevention and treatment of opportunistic infections, palliative care, and the elimination of stigma and discrimination remains crucial (Granich et al., 2017; Lau et al., 2011; Zigah et al., 2023). According to the JPR model's evaluation of the impact of interventions, local policies that are integrated into the regional situation can achieve better results. The implementation of local policies can make more effective use of financial and human resources, which has a positive impact on the control of the epidemic and can also provide an essential reference for the prevention and control of AIDS in other regions (Chen et al., 2019).

Moreover, blocking transmission pathways based on regional development patterns holds greater practical value. To illustrate,



(a)

(b)

(c)

Fig. 5. Raw incidence curves for different clustering results:(a) Type1: Southwest region; (b) Type2: Central and East region; (c)Type3: The other region.

consider the case of Xinjiang. Prevention and control of the epidemic in this region should be approached from the perspective of the development model, paying attention to stopping the spread of the AIDS epidemic caused by interregional drug transmission within China and intensifying efforts to study and control interprovincial drug transmission, which would be of greater value in terms of prevention and control. Lastly, Hotspot analysis reveals that the government should allocate more funds for prevention and control in the southwestern region, which remains the most severely affected area in China.

Despite its contributions, this study has certain limitations. Firstly, due to the unavailability of pre-2004 AIDS data in China's public health databases, analysis of earlier trends is not possible. Secondly, the study's scope is restricted to China's 31 provinces and municipalities, precluding examination of HIV trends and development patterns at smaller spatial units like counties, which could offer more granular insights. Future research should explore smaller areas to gain a comprehensive understanding of the epidemic's evolution.

6. Conclusion

This study employs a spatiotemporal analysis method based on the JPR model and K-means time series clustering, addressing the limitations in prior research regarding fine-grained temporal analysis and attempting to fill the gap in the area of spatial delineation of disease progression patterns. By examining the changes in China's AIDS epidemic from 2004 to 2018 in a multidimensional manner, the study assesses the impact of policies and risk factors on the epidemic, which further demonstrates the efficacy of the JPR model in intervention evaluation. Controlling sexual transmission routes, such as MSM, is a crucial direction for further AIDS prevention and control in China, with targeted regional policies proving more efficient. Zoning by development pattern aids in understanding the epidemic's spread and enhances control efforts. Additionally, this study utilizes hotspot analysis to identify high-risk and low-risk areas, helping to picture the epidemic's current status.

This study presents a fresh perspective on the development pattern of the AIDS epidemic and offers valuable insights and guidance for AIDS assessment and policymaking. Through understanding spatiotemporal development patterns, targeted recommendations can be made for AIDS prevention and control strategies.

CRediT authorship contribution statement

Shu-nan Gui: Writing – original draft, Visualization, Methodology,

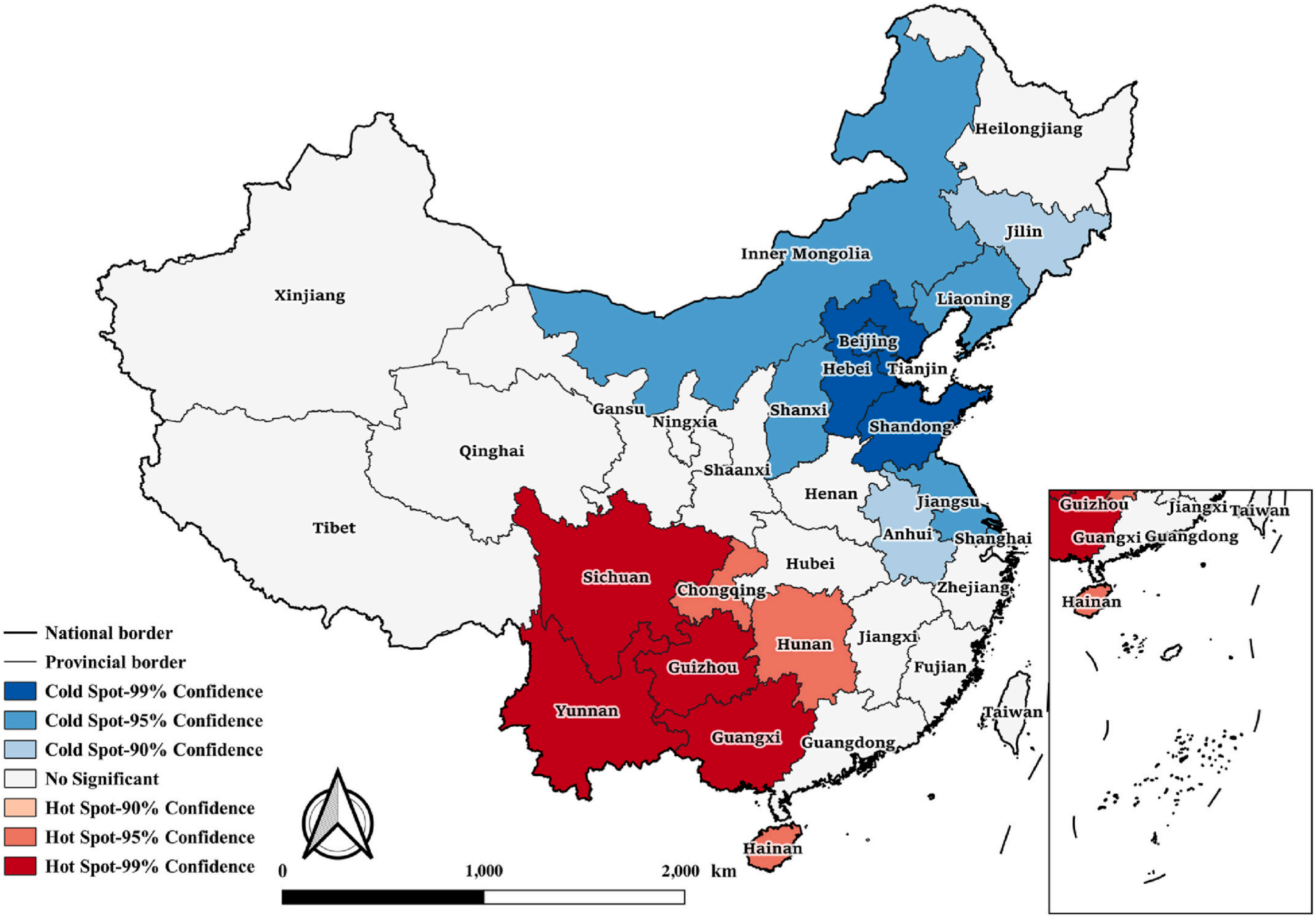


Fig. 6. Hotspot analysis result of AIDS incidence in China, 2018).

Investigation, Formal analysis, Data curation. **Xiang Zhang:** Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation. **Zhenhui Sun:** Writing – review & editing, Software, Methodology, Conceptualization. **Yao Yao:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Key Research and Development Program of China [2023YFB3906803]; the “CUG Scholar” Scientific Research Funds at China University of Geosciences [2022034], the Guangdong-Hong Kong-Macau Joint Laboratory Program [2020B1212030009] and the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources [KF-2023-08-04].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.healthplace.2024.103353.

References

Basavaraju, S.V., Mwangi, J., Nyamongo, J., et al., 2010. Reduced risk of transfusion-transmitted HIV in Kenya through centrally co-ordinated blood centres, stringent donor selection and effective p24 antigen-HIV antibody screening. Vox Sang. 99, 212–219.

Bonovas, S., Nikolopoulos, G., 2012. High-burden epidemics in Greece in the era of economic crisis. Early signs of a public health tragedy. J Prev Med Hyg 53, 169–171.

Box, G.E.P., Tiao, G.C., 1975. Intervention analysis with applications to economic and environmental problems. J. Am. Stat. Assoc. 70, 70–79.

Campbell, E.M., Jia, H., Shankar, A., et al., 2017. Detailed transmission network analysis of a large opiate-driven outbreak of HIV infection in the United States. J. Infect. Dis. 216, 1053–1062.

Carrel, M., Janko, M., Mwandagalirwa, M.K., et al., 2016. Changing spatial patterns and increasing rurality of HIV prevalence in the democratic republic of the Congo between 2007 and 2013. Health Place 39, 79–85.

Chen, Y., Tang, Z.Z., Tang, S., et al., 2018. Decreasing HIV, syphilis, and hepatitis c infection after a decade of harm reduction implementation among drug users in southwestern areas of China. J. Stud. Alcohol Drugs 79, 248–257.

Chen, H., Luo, L., Pan, S.W., et al., 2019. HIV epidemiology and prevention in southwestern China: trends from 1996-2017. Curr. HIV Res. 17, 85–93.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. Peerj computer science 7, e623.

1 State Council Office Of The People’s Republic Of China. Circular of the general office of the state council on the issuance of the 12th five-year plan of action for the containment and prevention of AIDS in China, 2012. https://www.gov.cn/gongbao/content/2012/content_2084242.htm. (Accessed 13 January 2012).

collaborators, G.H., 2019. Global, regional, and national incidence, prevalence, and mortality of HIV, 1980-2017, and forecasts to 2030, for 195 countries and territories:

a systematic analysis for the global burden of diseases, injuries, and risk factors study 2017. Lancet HIV 6, e831–e859.

Dou, Z., Chen, R.Y., Wang, Z., et al., 2010. HIV-infected former plasma donors in rural central China: from infection to survival outcomes, 1985-2008. PLoS One 5, e13737.

Feng, L.G., Ding, X.B., Lu, R.R., et al., 2009. High HIV prevalence detected in 2006 and 2007 among men who have sex with men in China's largest municipality: an alarming epidemic in chongqing, China. Jaids-Journal of Acquired Immune Deficiency Syndromes 52, 79–85.

Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. 24, 189–206.

Ghys, P.D., Williams, B.G., Over, M., et al., 2018. Epidemiological metrics and benchmarks for a transition in the HIV epidemic. PLoS Med. 15, e1002678.

Gokhale, R.H., Galang, R.R., Pitman, J.P., et al., 2017. A tale of 2 HIV outbreaks caused by unsafe injections in Cambodia and the United States, 2014-2015. Am. J. Infect. Control 45, 106–107.

Goovaerts, P., 2013. Analysis of geographical disparities in temporal trends of health outcomes using space-time joinpoint regression. Int. J. Appl. Earth Obs. Geoinf. 22, 75–85.

Granich, R., Williams, B., Montaner, J., et al., 2017. 90-90-90 and ending AIDS: necessary and feasible. Lancet 390, 341–343.

Gray, K.L., Walker, N.F., Martineau, F., et al., 2021. Interruption of tuberculosis detection and care during the ebola virus disease epidemic (2014-2015) in Liberia: time-series analyses for 2013-2017. Int. J. Infect. Dis. 112, 13–20.

Hamilton, J.D., 2020. Time Series Analysis. Princeton university press.

Hansen, B.E., 1995. Time Series Analysisjames D. Hamilton princeton university press, 1994.

He, N., 2007. Sociodemographic characteristics, sexual behavior, and HIV risks of rural-to-urban migrants in China. Biosci Trends 1, 72–80.

Hincapie-Castillo, J.M., Goodin, A., 2023. Using joinpoint regression for drug utilization research: tutorial and case study of prescription opioid use in the United States. Pharmacoepidemiol. Drug Saf. 32, 509–516.

Huang, J., Wu, H., Lin, S., et al., 2022. Spatial-temporal analysis of HIV/AIDS and syphilis in mainland China from 2007 to 2017. J. Med. Virol. 94, 3328–3337.

Jia, M., Luo, H., Ma, Y., et al., 2010. The HIV epidemic in yunnan province, China, 1989-2007. J. Acquir. Immune Defic. Syndr. 53 (Suppl. 1), S34–S40.

Jiang, J., Zhou, Y., Li, H., et al., 2018. HIV epidemic among pregnant women in China, 2016: trend and spatial analysis. Zhonghua Yixue Zazhi 98, 3360–3364.

Kalipeni, E., Zulu, L., 2008. Using gis to model and forecast HIV/AIDS rates in africa, 1986–2010. Prof. Geogr. 60, 33–53.

Kandwal, R., Garg, P.K., Garg, R.D., 2009. Health gis and HIV/AIDS studies: perspective and retrospective. J. Biomed. Inf. 42, 748–755.

Kim, H.J., Fay, M.P., Feuer, E.J., et al., 2000. Permutation tests for joinpoint regression with applications to cancer rates. Stat. Med. 19, 335–351.

Kulldorff, M., Athas, W.F., Feurer, E.J., et al., 1998. Evaluating cluster alarms: a space-time scan statistic and brain cancer in los alamos, New Mexico. Am. J. Publ. Health 88, 1377–1380.

Lau, J.T.F., Lin, C., Hao, C., et al., 2011. Public health challenges of the emerging HIV epidemic among men who have sex with men in China. Publ. Health 125, 260–265.

Li, M., Shen, Y., Jiang, X., et al., 2014. Clinical epidemiology of HIV/AIDS in China from 2004-2011. Biosci Trends 8, 52–58.

Li, L., Cuerden, M.S., Liu, B., et al., 2021. Three statistical approaches for assessment of intervention effects: a primer for practitioners. Risk Manag. Healthc. Pol. 14, 757–770.

Lin, X., Mao, X., Yao, W., 2023. Epidemiological characteristics and spatial clustering of intestinal infectious diseases in China, 2008—2018. Chinese General Practice 26, 417.

Ma, Y., Cui, Y., Hu, Q., et al., 2021. Long-term changes of HIV/AIDS incidence rate in China and the u.S. Population from 1994 to 2019: a join-point and age-period-cohort analysis. Front. Public Health 9, 652868.

Meng, X.J., Grulich, A., Wang, X.W., et al., 2018. Repeat HIV testing and incident rates among individuals attending voluntary counseling and testing clinics in wuxi, China: a retrospective study. Biomed. Environ. Sci. 31, 37–47.

Niu, D., Wang, K., Sun, L., et al., 2020. Short-term photovoltaic power generation forecasting based on random forest feature selection and ceemd: a case study. Appl. Soft Comput. 93, 106389.

Qian, S., Guo, W., Xing, J., et al., 2014. Diversity of HIV/AIDS epidemic in China: a result from hierarchical clustering analysis and spatial autocorrelation analysis. AIDS 28, 1805–1813.

Qiao, Y.C., Xu, Y., Jiang, D.X., et al., 2019. Epidemiological analyses of regional and age differences of HIV/AIDS prevalence in China, 2004-2016. Int. J. Infect. Dis. 81, 215–220.

General office of the standing committee of the Guangxi Zhuang autonomous region people's congress. Regulations on AIDS prevention and control in guangxi zhuang autonomous region, 2013. https://gxrd.gov.cn/html/art144870.html. (Accessed 6 April 2013).

Tao, Y., Chen, M.Y., Tucker, J.D., et al., 2020. A nationwide spatiotemporal analysis of syphilis over 21 years and implications for prevention and control in China. Clin. Infect. Dis. 70, 136–139.

Wang, L., 2007. Overview of the HIV/AIDS epidemic, scientific research and government responses in China. AIDS 21 (Suppl. 8), S3–S7.

Wang, X.Z., Smith, K., Hyndman, R., 2006. Characteristic-based clustering for time series data. Data Min. Knowl. Discov. 13, 335–364.

Wang, A., Qiao, Y., Dou, L., et al., 2018. Challenges of eliminating mother-to-child transmission of HIV, syphilis and hepatitis b in China: a cross-sectional survey. Lancet 392, S55.

Wang, R.Y., Feng, Z.X., Liu, Y., et al., 2020a. Is lifestyle a bridge between urbanization and overweight in China? Cities 99.

Wang, R.Y., Huang, B., Zhuolin, P., et al., 2020b. Spatial pattern evolvement of the distribution offoodborne illness in China and its influence mechanism. World Regional Studies 29, 168.

Wang, Y., Zhao, C., Liu, Z., et al., 2021. Spatiotemporal analysis of AIDS incidence and its influencing factors on the Chinese mainland, 2005-2017. Int. J. Environ. Res. Publ. Health 18.

Wei, Z., Zhijiang, Z., Lijun, W., 2017. Joinpoint regression analysis for the trends of cervical cancer mortality between 1987-2014 in China. China Oncology 27, 634–640.

Wong, J.A.H.A., 1979. Algorithm as 136: a k-means clustering algorithm. J. Roy. Stat. Soc. 28, 100–108.

Wu, Z., 2015. Achievement of HIV/AIDS Program in the Past 30 Years and Challenges in China.

Wu, Z., Chen, J., Scott, S.R., et al., 2019. History of the HIV epidemic in China. Curr. HIV AIDS Rep. 16, 458–466.

Xing, J., Guo, W., Qian, S., et al., 2014. Spatial analysis on the HIV/AIDS infection status among injecting drug users in China, from 2005 to 2011. Zhonghua liu Xing Bing xue za zhi= Zhonghua Liuxingbingxue Zazhi 35, 151–156.

Xu, S., Zhan, P., 2024. Twice-yearly lenacapavir demonstrated exceptional efficacy for HIV prevention. Innovation, 100683.

Xu, B., Li, J., Wang, M., 2020. Epidemiological and time series analysis on the incidence and death of AIDS and HIV in China. BMC Publ. Health 20, 1906.

Xu, J.J., Han, M.J., Jiang, Y.J., et al., 2021. Prevention and control of HIV/AIDS in China: lessons from the past three decades. Chin. Med. J. 134, 2799–2809.

Zeng, H., Chow, E.P., Zhao, Y., et al., 2016. Prevention of mother-to-child HIV transmission cascade in China: a systematic review and meta-analysis. Sex. Transm. Infect. 92, 116–123.

Zigah, E.Y., Abu-Ba'are, G.R., Shamrock, O.W., et al., 2023. "For my safety and wellbeing, i always travel to seek health care in a distant facility"—the role of place and stigma in HIV testing decisions among gbmsm–bsgh 002. Health Place 83, 103076.

Zulu, L.C., Kalipeni, E., Johannes, E., 2014. Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using gis: the case of Malawi, 1994-2010. BMC Infect. Dis. 14, 285.

PRC National Health And Family Planning Commission (NHFPC). Notice on further advancement of AIDS prevention and control, 2013. https://www.gov.cn/gzdt/2013-12/13/content_2547500.htm. (Accessed 13 December 2013).