Spatial autocorrelation and stratified heterogeneity in the evaluation of breast cancer risk inequity and socioeconomic factors analysis in China: Evidence from Nanchang, Jiangxi Province

Yaqi Wang,[1] Zhiwei Wan[2]

[1]Comprehensive Tumour Internal Department, Jiangxi Provincial Cancer Hospital, Nanchang; [2]School of Geography and Environmental Engineering, Gannan Normal University, Ganzhou, China

Abstract

Study of socioeconomic factors can play an important role in the spatial distribution of breast cancer by leading to a better understanding of its spatial pattern and assist breast cancer screening and early diagnosis. Taking Nanchang, a major city in central China, as an example, spatial autocorrelation and stratified heterogeneity were applied using a 10×10 km grid division to analyse breast cancer risk and socioeconomic factors. The research results showed that the median incidence rate of female breast cancer in Nanchang from 2016 to 2018 was 6.6/100,000 with a standard deviation of 12.3/100,000. Areas with higher incidence rates were mainly located in the central urban area and the major county towns. Spatial regression analysis showed that there was a statistically significant correlation between the spatial patterns of breast cancer incidence on the one hand, and on the other socioeconomic factors, such as total gross domestic product (GDP), per capita GDP and density of places of social and economic activities, *i.e.* points of interest. In addition, the normalized difference vegetation index also played a part in this respect. This research could serve as a reference for regional public health policy formulation and breast cancer screening.

Correspondence: Zhiwei Wan, School of Geography and Environmental Engineering, Gannan Normal University, Ganzhou 341000, China. E-mail: wzw3392008@sina.com

Key words: Female breast cancer; spatial autocorrelation; stratified heterogeneity; socioeconomic factors; risk detection; China.

Acknowledgements: we thank LetPub Accdon LLC for its linguistic assistance during the preparation of this manuscript. We thank anonymous reviewers and editors for their helpful comments and suggestions.

Funding: Gannan Normal University Doctoral Research Foundation (BSJJ202114) supported this study.

Received for publication: 1 February 2022.
Revision received: 29 April 2022.
Accepted for publication: 29 April 2022.

Geospatial Health 2022; 17:1078
doi:10.4081/gh.2022.1078

Introduction

Breast cancer is one of the most common cancers across the world. It is the second leading cause of death in developed countries after lung cancer, accounting for more than 30% of newly discovered cancers in women. Statistics from the International Agency for Research on Cancer (IARC) show that among the 19.3 million new cancer patients in the year 2020, female breast cancer has become the most common cancer, with an estimated 2.3 million new cases accounting for nearly 12% of total cases (Ferlay *et al*., 2021; Sung *et al*., 2021). In 2020, about 685,000 people died from breast cancer, which ranked first among all female cancers (Lei *et al*., 2021). Data from the United States also show that the number of breast cancers in women accounted for 29% of all cancers in 2015 (Siegel *et al*., 2015), which represent a serious burden on women's health.

Through China's economic reform and opening-up policy, its economy and society have developed rapidly. However, during this period, the mortality rate of breast cancer in China increased by more than 90%. Breast cancer is one of the most malignant tumours affecting women (Li *et al*., 2011; Fan *et al*., 2014; Chen *et al*., 2018). Global tumour epidemic statistics (GLOBOCAN) show that in 2018 China added about 200,000 cases of breast cancer and about 50,000 people died of the disease (Bray *et al*., 2018). Genetics, the immediate environment, lifestyle and socioeconomic factors all influence its incidence and treatment (Seidman *et al*., 1982), while general screening can increase early diagnosis (Wang *et al*., 2012). Many recent studies have shown that social factors affect the incidence of breast cancer (Ghoncheh *et al*., 2016) and that they are important from an epidemiological point of view. The incidence of breast cancer in developed countries is higher than in the developing countries (Igene, 2008) and, according to China's statistics (Chen and Zheng, 2015; Zhou *et al*., 2018), the incidence of breast cancer (34.3 per 100,000 women) in urban areas during the past 30 years is significantly higher than that in the rural areas

(17.0 per 100,000 women). Socioeconomic factors have come into focus in the epidemiological analysis of this cancer (Lantz *et al*., 2006; Ren *et al*., 2019; Ji *et al*., 2020) strengthening the role of risk identification and influencing the study of the mechanisms involved.

In recent years, with the development of geographic information system (GIS) technology (Macquillan *et al*., 2017; Ramadan *et al*., 2017), various spatial analysis methods have been used in disease risk detection (Ge *et al*., 2016; Xu *et al*., 2019). In particular, GIS and remote sensing (RS) have been increasingly used in the research of the transmission of infectious diseases (Qi and Du, 2013), as well as for the spatial distribution of chronic diseases in relation to social equity (Wang *et al*., 2010) and the spatial and temporal matching of medical and health resources to the presence of disease (Salinas *et al*., 2018). These techniques have been especially useful in the study of public health emergencies, such as those caused by severe acute respiratory syndromes caused by different corona viruses (Wang *et al*., 2006; Fan *et al*., 2020). Since many tumour epidemic factors have certain geospatial characteristics (Al-Ahmadi and Al-Zahrani, 2013; Xia *et al*., 2016), researchers have begun to use spatial autocorrelation analysis for spatial recognition in areas with high disease incidence (Sun *et al*., 2017). Stratified heterogeneity detection technology can then be employed to identify differences in disease incidence caused by different social and economic factors (Liao *et al*., 2019).

GIS technology is not widely used in breast cancer research in China. Most existing studies are based on the data from one province in China to study the spatial pattern of breast cancer incidence (Chu *et al*., 2017), and few studies have focused on breast cancer incidence in metropolitan areas or analysed incidence differences between urban and rural areas in China. To address this research gap, we selected the city of Nanchang in central China as research area since this city is the capital of Jiangxi Province and has a mid-position with respect to gross domestic product (GDP) and population, *i.e*. it ranks as number 15 among mainland China's 31 provinces (Qi *et al*., 2013; Liu *et al*., 2016). Therefore, the research in this area has a certain representativeness, and the Poyang Lake Plain where Nanchang is located is large meaning that research here would include both a concentrated and a dispersed population profile, thus representing a typical situation of urban and rural areas. In this study, spatial autocorrelation and stratified heterogeneity detection were used to examine the incidence of female breast cancer and its relationship between different levels of socioeconomic factors.

Materials and methods

Research area

Nanchang, situated close to the main tributary of the middle and lower reaches of the Yangtze River between 115°27'E to 116°35'E and 28°10'N to 29°11', is the centre of the Poyang Lake Ecological Economic Zone and one of the major cities in the Yangtze River Economic Belt (Figure 1). Its central and eastern parts are plains, while the western part includes hills and more

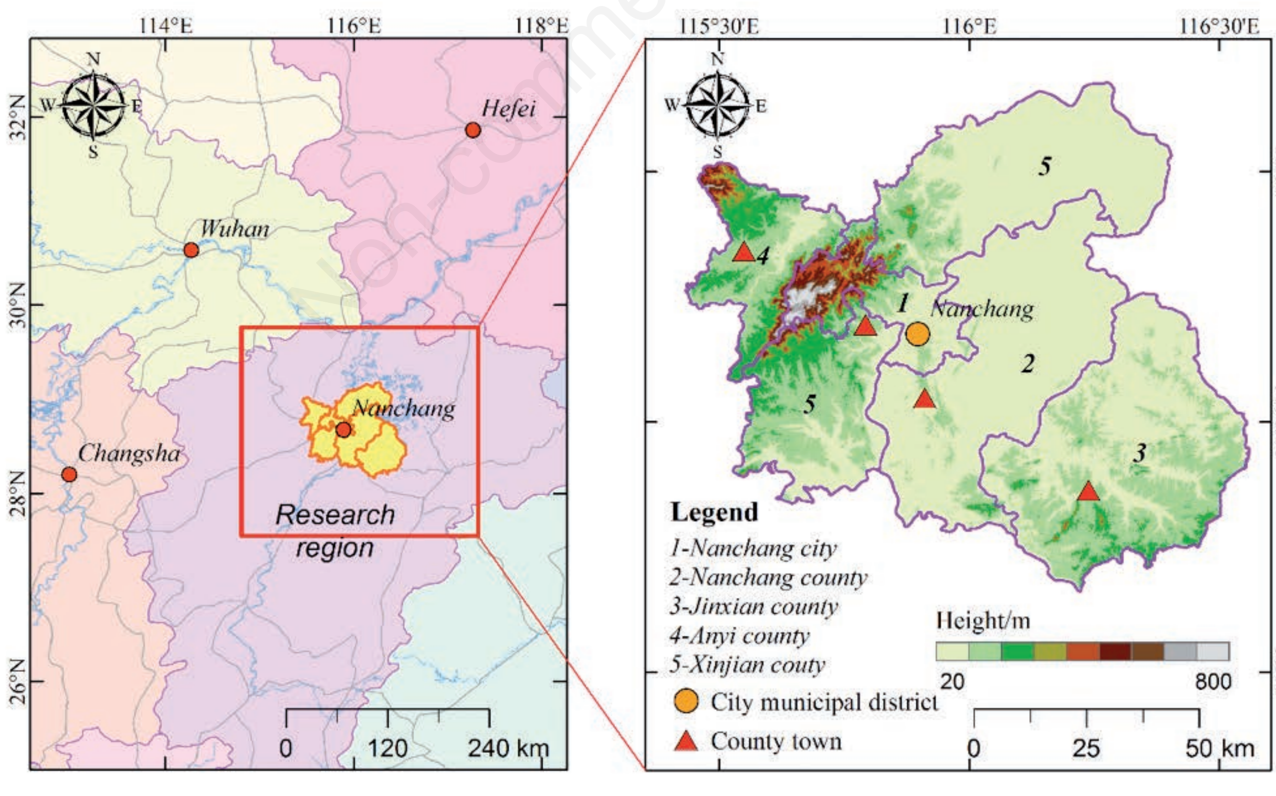

Figure 1. Map and administrative divisions of the study area.

mountainous areas. Administratively, there is an urban area and four counties, with a total area of 7402 km^2 and a population of approximately 5.5 million. Data from the sixth national census show that the urbanized part of Nanchang is about 50% of the total. In terms of physical geography, Nanchang has a subtropical monsoon climate with suitable precipitation and heat conditions for its well-developed agriculture. The Poyang Lake Plain is China's main commercial rice production base. The urban areas of the study area are mainly concentrated in the municipal district of Nanchang City and its four county towns. The remaining areas are mainly market towns and rural areas.

Medical record data and spatial processing

This research collected the medical records of all breast cancer patients, including patient address information and diagnosis information after privacy information processing, in Nanchang in Jiangxi Province Cancer Hospital from 2016 to 2018. Due to the random distribution of case data, previous studies have tried to obtain a continuous incidence surface through regional kernel density estimation and Kriging estimation (Kloog *et al*., 2009) as this approach effectively eliminates the ecological fallacy and modifiable areal unit problem (MAUP) of case data (Nelson and Brewer, 2017). We used Microsoft Access 2003 database and ArcGIS 10.2 software (ESRI; Redlands, CA, USA) for database construction and spatial information management, respectively. We used the Baidu map address space matching interface (http://api.map.baidu.com/) to spatially identify the addresses as described by Tan *et al*. (2021) thereby achieving precise spatial positioning of the cancer patients through the geographical coordinates of their residential addresses. Cases with missing address and diagnosis information were not counted.

The population data came from China's sixth national population census data. The GIS data for administrative divisions were obtained from the Geospatial Data Cloud of the Chinese Academy of Sciences (http://www.gscloud.cn/). Using the WGS-1984 coordinate system, we divided the study area into a grid with 10×10 km cells (Figure 2) for the reason that Nanchang includes 109 township administrative units, which is approximately equal to the 10-km cell size.

Socioeconomic data

We considered the availability and representativeness of the data and referred to the existing literature to select the following six indicators as social and economic factors for analysis (Figure 3):

- Population density, derived from China's km-grid population distribution dataset (Fu *et al*., 2014). This dataset spatialized the data of China's sixth population census from 2010 to derive high-resolution population spatial distribution data. The relevant data were rigorously corrected and tested and considered representative;
- GDP, derived from China's km grid GDP distribution dataset (Huang *et al*., 2014). This dataset provides nationwide km-resolution GDP spatial distribution by high-precision grid interpolation based on the 2010 Chinese GDP and applies strict data correction and inspection;
- GDP per capita was calculated using the grid calculator tool of

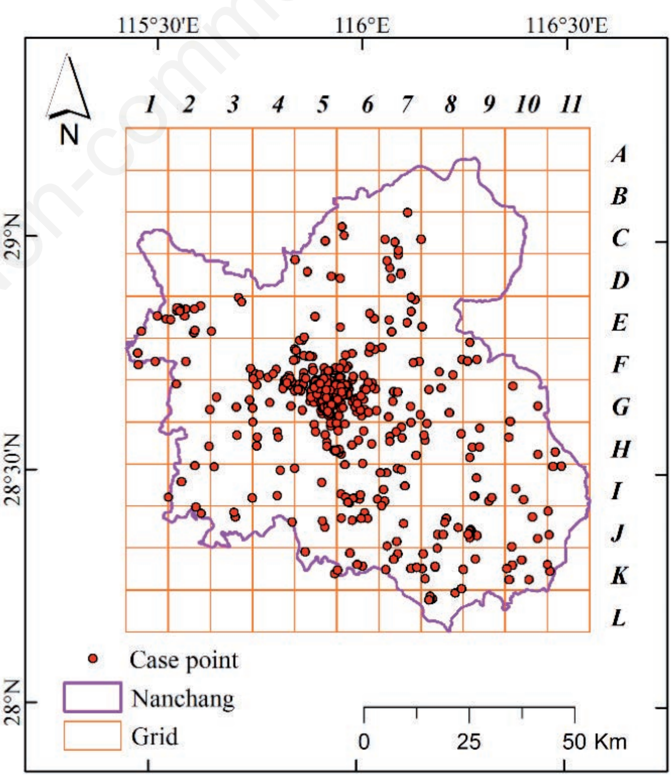

Figure 2. Spatial distribution and grid division of female breast cancer cases in Nanchang 2016-2018.

ArcGIS software using the above stated population (1) and GDP (2) datasets;
- The density of places of various social and economic activities in the area, *i.e*. points of interest (POIs) The number of POIs per km-grid cells represents the density of the socioeconomic activities and is an important indicator of the state of social and economic development. The POI data used in this research came from the Baidu map API (http://api.map.baidu.com/place);
- Night-time light data, which indicate the economic activity. The higher the night light index, the more developed the economic activity in the area. The night-time light data used in this research came from the DMSP/OLS at the National Geophysical Data Centre (https://www.ngdc.noaa.gov/eog/dmsp/downloadV4 composites.html); and
- The normalized difference vegetation index (NDVI), which reflects the regional vegetation status. Generally, the closer the NDVI is to 1, the greener the surface. NDVI <0 generally represents water bodies and these areas were therefore not included in our calculations. The NDVI data came from the Geospatial Data Cloud of the Chinese Academy of Sciences (http://www.gscloud.cn/) and represent the annual average of Landsat satellite imagery from 2010 calculated by the Environment for Visualizing Images (ENVI) software, version 5.1.

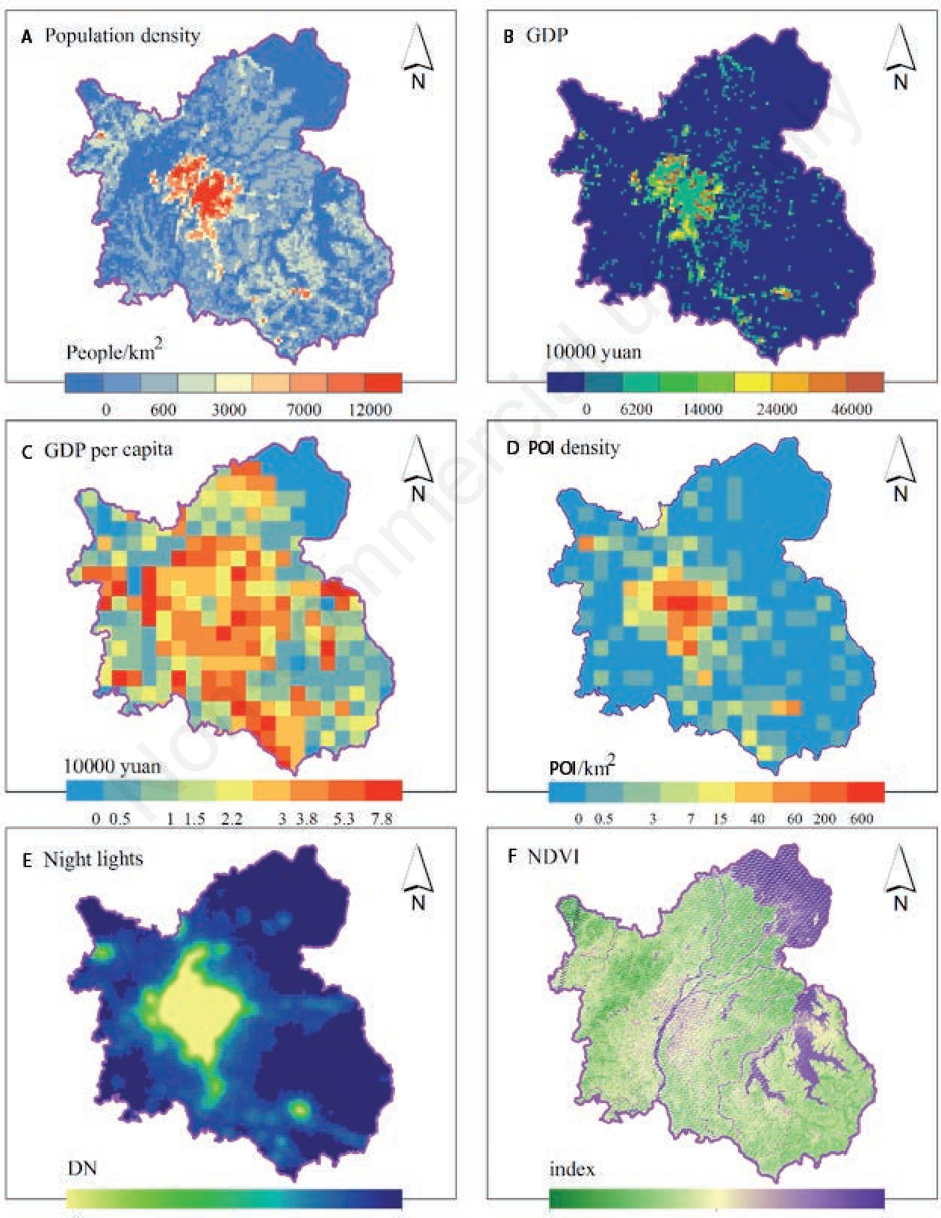

Figure 3. The spatial distribution map of the main social and economic factors in Nanchang. A) Population density; B) gross domestic product (GDP); C) GDP per capita; D) points of interest (POI) density; E) night lights; F) normalized difference vegetation index (NDVI).

Research methods

Pearson correlation analysis

Pearson correlation (r) analysis was used to calculate the relationship between various socioeconomic factors and the incidence of tumours, and its significance. In general, r ranges from –1 to +1; r=1 means that the two factors are positively related, r= –1 means that the two factors are negatively related and r=0 means that the two factors are unrelated. The calculation formula of the correlation coefficient is:

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (1)$$

where $\bar{x}$ and $\bar{y}$ are the average values of the factors investigated.

Spatial autocorrelation

The spatial pattern of disease events can be identified by Global Moran's I, which also varies between –1 and +1. The closer to –1, the higher the negative correlation of the spatial distribution of the factors; the closer to +1, the higher the positive correlation (Liao *et al*., 2010; Wang *et al*., 2012). The relevant formula is:

$$\text{Moran's } I=\frac{\sum_{ij}(x_i-\bar{x})(x_j-\bar{x})}{S^2\sum_{ij}^{n}w_{ij}} \quad (2)$$

where x_i and x_j are sample values; $\bar{x}$ the mean; w_{ij} the spatial weight matrix; and S^2 the mean square error.

Local Moran's I (Anselin, 1995) reflects the aggregation status and pattern of individual spatial units better, so it is often used to identify hotspots and coldspots, which are signs of areas characterized by abnormal disease incidence. The formula is:

$$\text{Local Moran's } I=\frac{n(x_i-\bar{x})\sum_{j}w_{ij}(x_j-\bar{x})}{\sum_{j}(x_i-\bar{x})} \quad (3)$$

where n is the number of grid cells; w_{ij} the spatial weight matrix; x_i and x_j sample values; and $\bar{x}$ the mean.

Stratified heterogeneity detection

Heterogeneity can be categorized into spatial local heterogeneity (SLH) and spatial stratified heterogeneity (SSH) (Graif and Sampson, 2009; Wang *et al*., 2016). SLH shows that there is a significant difference between the attribute values of a certain area and the surrounding area, which can help identify an abnormal area (Chung *et al*., 2014). SLH is determined by local spatial autocorrelation and other methods. SSH generally means that there is a significant difference in certain attributes between different types in the region. For example, regions with different socioeconomic conditions have significant differences in the incidence of a certain disease. In the past, owing to lack of quantitative measurement methods, SLH was generally described qualitatively (Wang *et al*., 2012). In recent years, geographic detector technology has proved to be an effective method for identifying spatial layering heterogeneity (Shrestha and Luo, 2018; Xu *et al*., 2018). Geographic detector technology is a stratified heterogeneity recognition method proposed by Wang *et al.* (2010) for studying the mechanisms influencing the division of space or social factors with regard to the risk of disease incidence. The crux of this method is to use the q value to quantify the degree at which the influencing factors (such as a certain natural or social factor) explain the spatial heterogeneity of the target factors (such as the incidence of a certain disease). The value range of q is between 0 and 1; q=1 means that the influencing factor fully explains (100%) the spatial heterogeneity of the target factor. The calculation of q *is* as follows:

$$q_{D,H}=1-\frac{1}{n\sigma_H^2}\sum_{i=1}^{m}(n_{D,i}\cdot\sigma_{H_{D,i}}^2) \quad (6)$$

Where $n_{D,i}$ represents the sample size of the risk factor D in the subregion i; n all sample sizes in the entire study area; and σ_H^2 the discrete variance of the area. When the risk factor D is decisive for the target factor (such as a certain type of disease), the discrete variance $\sigma_{H_{D,i}}^2$ of each sub-region is small and the variance between them large. In extreme cases, when $\sigma_{H_{D,i}}^2\rightarrow 0$ and, $q_{D,H}$=1 the spatial distribution of the disease is completely determined by this factor.

Spatial lag model and spatial error model

Ordinary least squares (OLS) is a linear regression method used to model the relationship between the incidence rate and various related elements. The OLS model assumes that the variables are spatially independent of each other (Wang *et al*., 2013; Fan *et al*., 2017). However, the spatial distribution of many social and economic factors and diseases has a certain spatial correlation. For example, surrounding communities of communities with higher incidence rates also have higher incidence rates. Therefore, we used the spatial econometrics methods of the spatial lag model (SLM) and the spatial error model (SEM) to solve this problem (Elhorst, 2010; Debarsy *et al*., 2018). In this study, the GeoDa-1.12.1.59 software (https://spatial.uchicago.edu/geoda) was used to complete the calculations of the spatial lag model and the spatial error model.

Other statistics used

Through Local Moran's I, all grids are divided into H-H groups indicating cancer incidence hotspots and other groups. The independent sample t-test in SPSS 20 software (https://www.ibm.com/analytics/spss-statistics-software) was used to test the incidence of breast cancer in these two groups of grid cells.

The Levene test is an equal variance test which can be used to check if data sets conform the homogeneity of variance assumption before the t-test (Nordstokke and Zumbo, 2010). The Breusch-Pagan test is an effective method used to determine whether heteroscedasticity is present in a regression model or not (Halunga *et al*., 2017). The Akaike information criterion (AIC) is an estimator of prediction error, *i.e.* it evaluates how well a model fits the data it was generated from. AIC is used to compare different possible models and determine which one is the best fit for the data.

Results

Based on the statistics of the incidence of female breast cancer in the study area from 2016 to 2018, a spatial distribution map of the average incidence of breast cancer in the study area was obtained for the study period (Figure 4). Across the 102 grid cells, the highest incidence rate was 73.2/100,000, while the lowest was 0; the median value was 6.6/100,000 with a standard deviation (SD) of 12.3/100,000. The higher-incidence grid cells were mainly located in urban central areas, such as G5, G6, H6, and J9 corresponding to the urban areas of Nanchang, Xinjian County, Nanchang County and Jinxian County, respectively. Grid cells with a low incidence rate generally had small populations and were located at the fringes of the study area, especially the Poyang Lake area in the northeast of the study area.

Spatial autocorrelation and identification of high-risk areas

The results of the global spatial autocorrelation analysis show that the incidence of female breast cancer in each grid cell in Nanchang had a spatially significant spatial correlation (Figure 5). The Geoda calculation results showed that the global Moran's *I* value was equal to 0.280 (P=0.0001) indicating that breast cancer had a high incidence rate, with a similar situation in the surrounding grid cells, while grid cells with a low incidence rate were more likely to have surrounding grid cells characterized by lower incidence rates of breast cancer. The results of the local spatial autocorrelation analysis showed that grid cells G6 and H6 belonged to the High-High (H-H) type (P=0.001) (Figure 6). The remaining H-H type grids were distributed in the centre of the study area and passed the confidence test with P=0.05. The Low-Low (L-L) type of breast cancer incidence grid cells were all distributed in the northwest and northeast areas of Nanchang and most of these areas belonged to economically underdeveloped areas such as mountainous and lake regions.

Correlation between incidence and socioeconomic factors

Through regression analysis of the incidence of female breast cancer and the social and economic factors, we found that the incidence showed a significant positive relationship with population density, total GDP, GDP per capita, POI density, night-time light index and other factors (Figure 7). These results indicate that grid cells with higher incidence of breast cancer also have a higher population density, which has a direct relationship with the degree of economic development. Breast cancer incidence also has a significant positive correlation with indicators that directly represent economic factors, such as the total GDP and per capita GDP (Table 1). As GDP and GDP per capita increase, the incidence of breast cancer in the grid also increases. The POI density and night-time light index also reflect the economic and social development of the region. The results of the regression analysis showed that the correlation coefficients of breast cancer incidence and POI density and night-time light index were 0.602 and 0.524, respectively. However, the incidence of breast cancer and NDVI showed an insignificant negative correlation (–0.079). Although not statistically significant (P=0.488) the result indicates that the better the

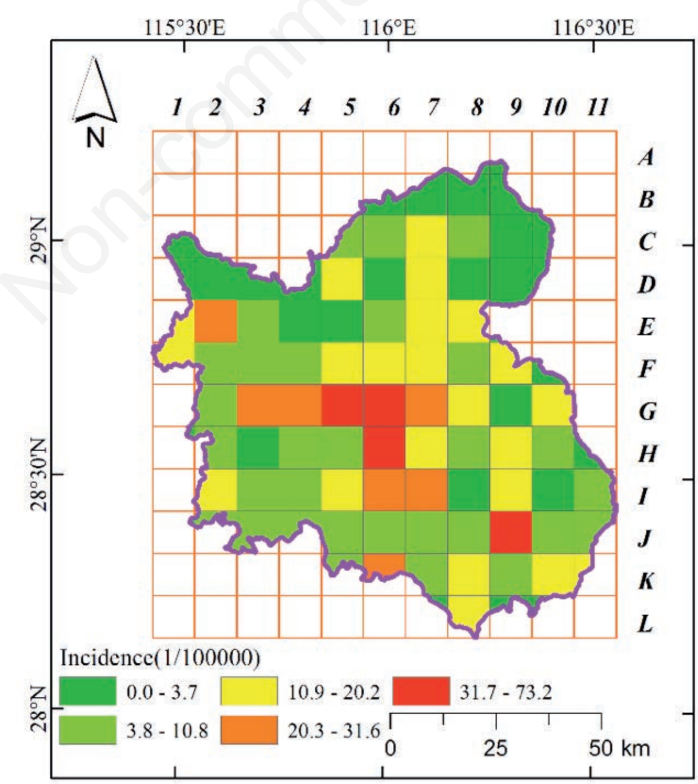

Figure 4. Spatial distribution of the average incidence of female breast cancer in Nanchang from 2016 to 2018.

regional vegetation status, the lower the incidence of breast cancer. This may be related to the low population density in the suburbs and slower socioeconomic development.

Variance analysis of the incidence rates in high-risk areas and low-risk areas

In order to better compare the spatial distribution of the incidence of breast cancer in Nanchang, based on the calculation results of local spatial autocorrelation, all the grid cells in the study area were sorted into the different correlation types: H-H, L-L, H-L and L-H. The H-H type that represents hotspots of female breast cancer were seen in a total of 10 grid cells (F5-7, G4-7, H4, H6-7). The independent sample t-test in SPSS 20 software was used to test the incidence of breast cancer in these two groups of grid cells (the hotspot cells and the others). The results are shown in Table 2. Under the assumption that the variances of the two groups were equal, we found t=6.666 (P<0.001); under the assumption that the variances of the two groups were not equal it was clearly different, t=3.617 (P=0.005). Therefore, it was considered that there is a significant difference in the incidence of breast cancer between the two groups of grid cells.

Spatial regression analysis

Taking the incidence of breast cancer in Nanchang as the dependent variable and population density, total GDP, GDP per capita, POI density, night-time light index and NDVI as independent variables, the OLS, SLM and SEM models were used to analyse the influencing factors of the incidence of breast cancer in Nanchang. As shown in Table 3, three conclusions could be drawn: i) SLM and SEM were superior compared to OLS from the perspective of goodness of fit (r^2) indicating that they, after considering the spatial autocorrelation effect, better reflected the incidence of breast cancer in the region; ii) SLM and SEM models produced lower numbers than OLS in terms of the AIC, which indicated that these methods were more effective than OLS; iii) The social and economic variables, except for population density and night-time light index, passed the significance test at P≤0.1. This shows that the total GDP, GDP per capita, POI density and NDVI index had an influence on the spatial distribution pattern of female breast cancer. Since Nanchang is a relatively developed area in Jiangxi Province, its population distribution has a certain density even in non-urban core areas. The rural areas in the study area also have certain economic activities and good transportation and power infrastructure. Therefore, the failure of the night-time light index to pass the significance test may be due to the relatively small differences between urban and rural areas.

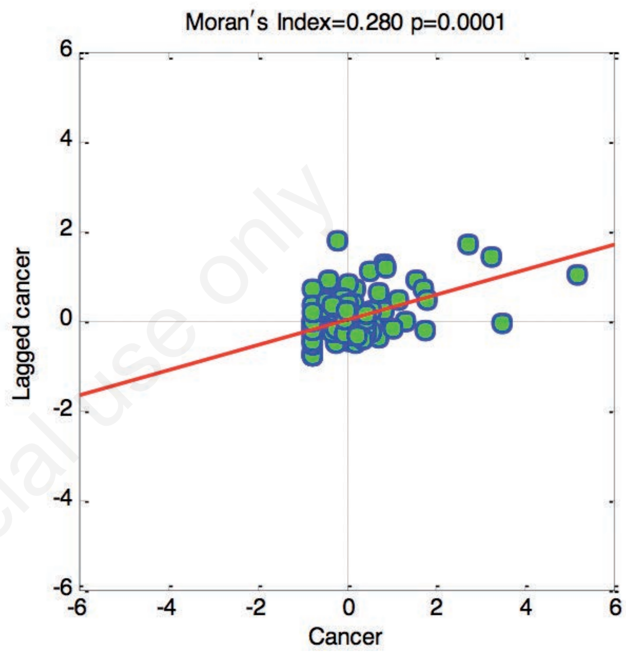

Figure 5. Global spatial autocorrelation Moran's *I* map of female breast cancer incidence in Nanchang.

Table 1. Regression results of breast cancer incidence and socioeconomic factors in Nanchang.

No.	Factor	Model	*r*	r^2	F-statistics	P-value
1	Population density	y=0.007x+5.852	0.603	0.364	57.240	<0.001
2	GDP	y=3E-05x+6.315	0.557	0.311	45.252	<0.001
3	Per capita GDP	y=2.800x+4.718	0.337	0.114	12.941	<0.001
4	POI density	y=0.212x+8.233	0.602	0.363	57.142	<0.001
5	Night-time light	y=0.550x+5.650	0.524	0.274	37.896	<0.001
6	NDVI	y=–9.762x+12.705	–0.079	0.006	0.486	=0.488

r, Pearson correlation; r^2, coefficient of determination; F-statistics, Fisher's statistics; GDP, gross domestic product; POI, points of interest; NDVI, normalized difference vegetation index.

Table 2. The incidence rate of female breast cancer in Nanchang calculated by the t-test.

	Levene test F-statistics	P	t	df	P	t-test Mean difference	Std. error difference	95% confidence interval Lower	Upper
Equal variances assumed	14.554	0.000	6.666	100	0.000	22.969	3.445	16.132	29.805
Equal variances not assumed	-	-	3.617	9.394	0.005	22.968	6.350	8.694	37.242

F-statistics, Fisher's statistics; Std. error, standard error; df, degrees of freedom.

Regional incidence heterogeneity under stratification of socioeconomic factors

Through the ArcGIS 10.2 software, the value of the various social and economic factors over the entire grid were divided into three levels (high, medium and low) according to equidistance classification. Using GeoDetector_2015 software (http://www.geodetector.cn/), the statistics of breast cancer incidence rate under different levels of socioeconomic factors were calculated. The results of the stratified heterogeneity analysis are shown in Table 4. The population density, total GDP, GDP per capita, POI density and q statistics of NDVI were between 0.135 and 0.298, with the P-values <0.001. This shows that these five socioeconomic factors can impact the spatial pattern of breast cancer incidence in Nanchang. However, the stratification status of the night-time light index had little effect on the spatial distribution pattern of breast cancer incidence in Nanchang, with a q value of only 0.024 and a less impressive level of significance (P>0.1).

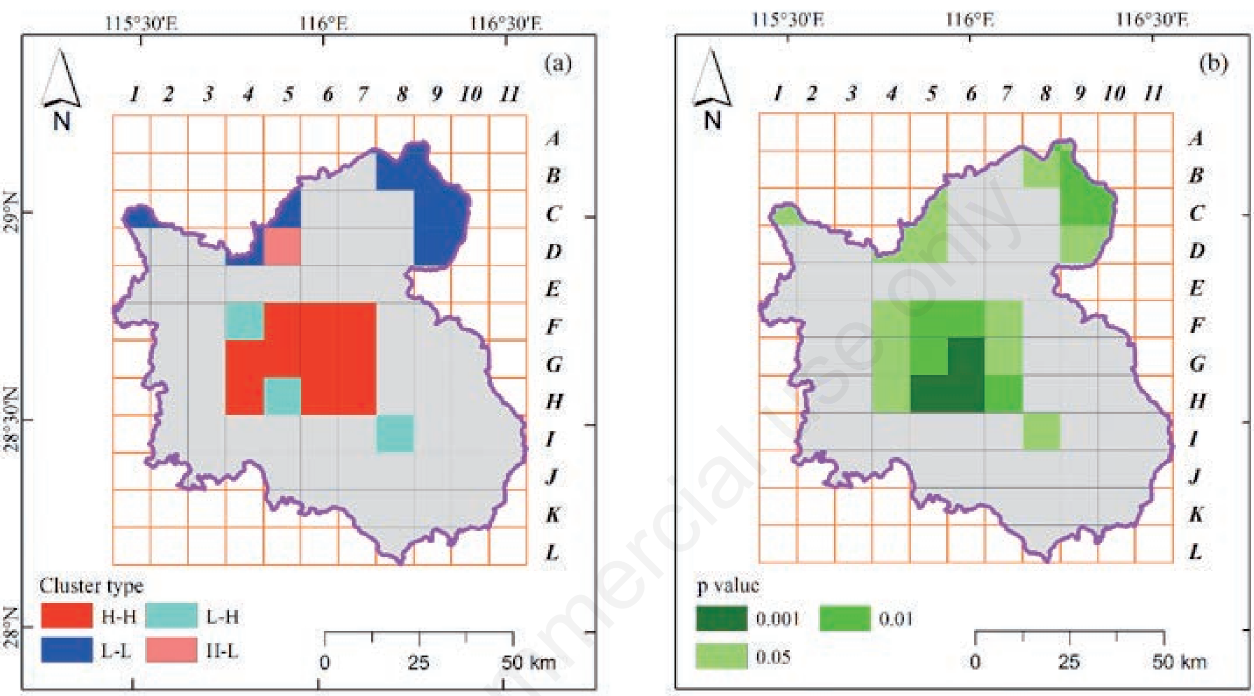

Figure 6. Local spatial autocorrelation types (A) and P-values of the incidence of female breast cancer (B) in Nanchang.

Table 3. Results for the spatial autoregressive models.

Variable	OLS			SLM			SEM		
	Coefficient	Std. error	t	Coefficient	Std. error	z	Coefficient	Std. error	z
Weight matrix	-	-	-	0.248**	0.114	2.168	0.208*	0.122	1.707
Constant	5.254**	1.601	3.280	3.646**	1.626	2.224	4.973***	1.726	2.879
Population density	–0.007	0.004	–1.605	–0.005	0.004	–1.378	–0.005	0.004	–1.304
Total GDP	3.9E-05**	1.5E-05	2.468	3.0E-05**	1.5E-05	2.004	2.9E-05**	1.5E-05	1.996
Per capita GDP	1.273*	0.751	1.695	1.173*	0.707	1.659	1.299*	0.742	1.749
POI density	0.280***	0.069	4.003	0.263***	0.065	4.013	0.260***	0.067	3.860
Night-time light	–0.252	0.179	–1.405	–0.253	0.170	–1.490	–0.178	0.184	–0.967
NDVI	13.814**	4.857	2.843	11.076**	4.711	2.350	13.738*	5.350	2.567
Test									
r^2	0.512			0.539			0.528		
Breusch-Pagan test	28.059			27.494			27.200		
logL	–364.168			–362.180			–363.183		
AIC	742.336			740.359			740.366		

OLS, ordinary linear regression; SLM, spatial lag model; SEM, spatial error model; Std. error, standard error; GDP, gross domestic product; POI, points of interest; NDVI, normalized difference vegetation index; logL, Log Likelihood; AIC, Akaike information criterion. *Significant at the 0.1 significance level; **significant at the 0.05 significance level; ***significant at the 0.01 significance level.

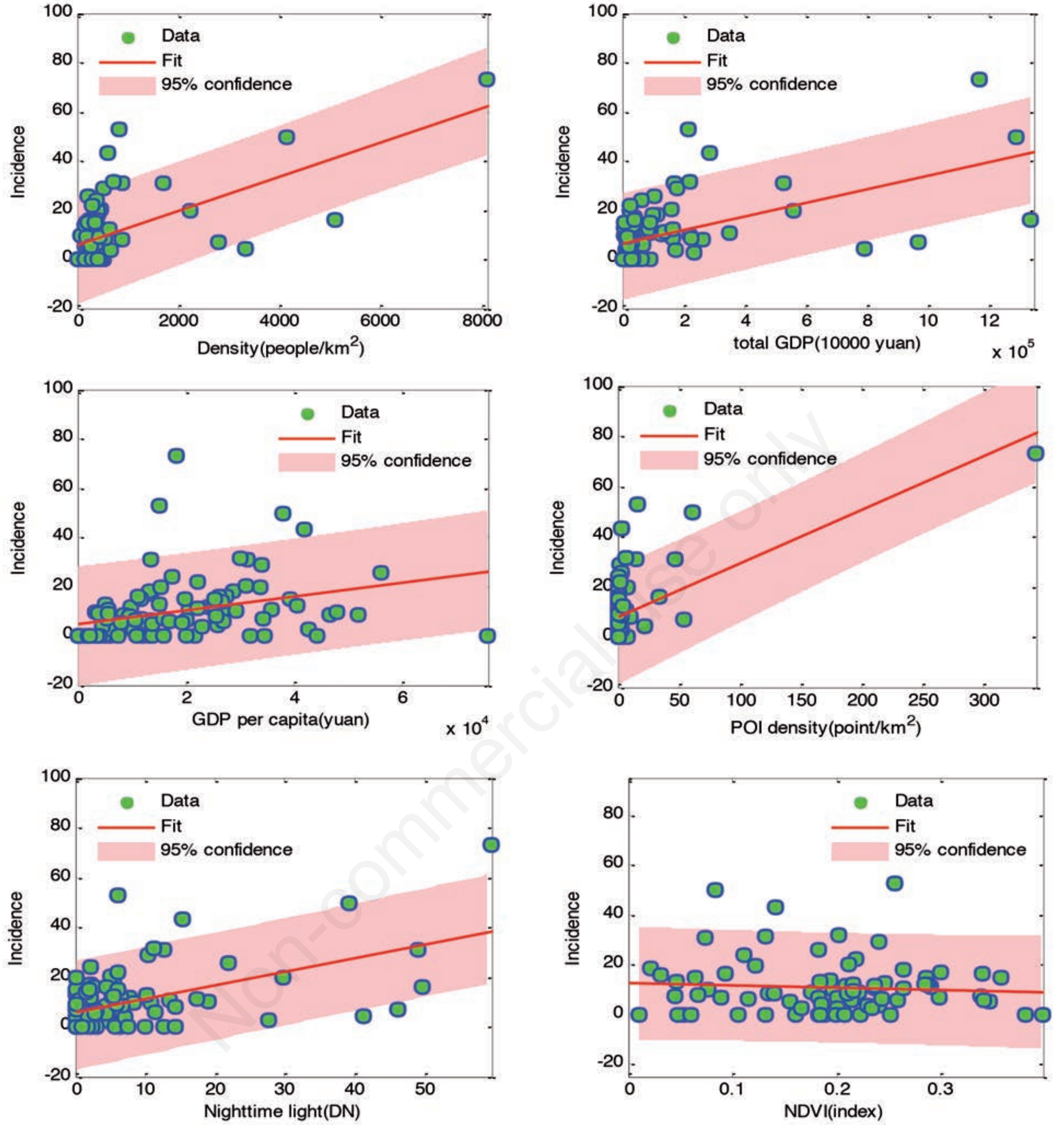

Figure 7. Regression analysis of female breast cancer incidence and socioeconomic factors in Nanchang. GDP, gross domestic product; POI, points of interest; NDVI, normalized difference vegetation index.

Table 4. Stratified heterogeneity test of female breast cancer incidence and socioeconomic factors in Nanchang.

Statistic	Population density	Total GDP	GDP per capita	POI density	Night-time light	NDVI
q statistic	0.186	0.298	0.175	0.270	0.024	0.135
P-value	<0.001	<0.001	<0.001	<0.001	0.307	<0.001

GDP, gross domestic product; POI, points of interest; NDVI, normalized difference vegetation index.

Discussion

The data of breast cancer cases selected in this study come from the largest oncology hospital in Jiangxi Province, and the hospital is located in Nanchang City. Due to the convenience of local treatment insurance reimbursement, most patients would choose to receive treatment locally. Therefore, the data sources have a high degree of consistency. Although this study is not the result of a comprehensive census, given China's high health insurance coverage, the proportions of all levels of society represented by the study cases are reliable. In future studies, more detailed cancer census and sample survey data can be attempted to further analyse the relationship between breast cancer incidence and socioeconomic factors. On the other hand, due to the development of China's economy and the increase in the proportion of transient population, it is inevitable that many registered addresses are only temporary addresses. It is also possible that even if the registered address is the correct location of the patient's household registration, the time of residence at this address can be very short. When discussing the influence of socioeconomic factors on the spatial distribution pattern of breast cancer incidence, patient long-term residency is required. In future research, we intend to pay great attention to this issue, and the improvement of research reliability requires more detailed patient survey and follow-up information.

Due to the uneven regional development in China, there is a great variation in regional development and therefore the incidence of breast cancer varies somewhat from region to region. Nevertheless, the overall differences in socio-economic development, particularly between urban and rural areas, can have an impact on regional patterns of breast cancer distribution. Studies have shown that there are great differences in the incidence and mortality of breast cancer in different regions of China. Xia *et al.* (2016) conducted a spatiotemporal analysis of geographic differences in distribution based on breast cancer data from retrospective surveys from 1973-1975, 1990-1992 and 2004-2005 using data from the Chinese Ministry of Health, showing that there were significant differences and a tendency for geographic differences to widen over time.

The choice of study area can have an impact on the reliability of the study findings. The choice of study area was based on a combination of data availability, regional representativeness and an appropriate urban-rural development gradient. Despite being the largest city in Jiangxi Province, the rural areas under Nanchang's jurisdiction have a huge disparity in economic and social development with the urban centres due to a considerable disparity in urban and rural development in China. Therefore, not all of Nanchang was selected for this study as a developed urban area, but also included a large proportion of rural areas.

Due to the uneven development of China's regions, although Jiangxi has a certain representativeness in the country, more detailed data are needed to study the spatial pattern of breast cancer incidence in other regions. Nevertheless, the influence of urban and rural socioeconomic differences on the spatial distribution pattern of breast cancer incidence revealed in this paper will provide certain theoretical support for the formulation of regional cancer control policies in the future. The current study focuses on socioeconomic factors and their relationship with the incidence of breast cancer. Natural factors such as geomorphological, climatic and hydrological factors are also important factors influencing the spatial distribution patterns of the disease and in the future, we expect to collect more data and at a higher resolution to further analyse the possible associations between these factors and the distribution patterns of breast cancer.

Conclusions

The incidence of breast cancer varies substantially under different socioeconomic development conditions. Therefore, clarifying the relationship between relevant factors and breast cancer incidence will play an important role in the future screening and early diagnosis and treatment of breast cancer. The use of spatial analysis technology to identify and detect hotspots of breast cancer is conducive to improving the efficiency and accuracy of breast cancer screening in the future. This study focused on Nanchang, a major city in central China. Based on the grid division, spatial pattern, and spatial autocorrelation characteristics, the high-risk areas of female breast cancer incidence from 2016 to 2018 were identified. Using the spatial stratification heterogeneity detection method, we systematically studied the association of different socioeconomic development factors with the incidence of female breast cancer.

Our research provides technical support for regional cancer screening and cancer prevention policy formulation and can be a reference for understanding regional risk factors for cancer. This paper also provides a basis for improving cancer risk screening efficiency and selection of screening strategies.

References

Al-Ahmadi K, Al-Zahrani A, 2013. Spatial autocorrelation of cancer incidence in Saudi Arabia. Int J Env Res Pub He 10:7207-28.

Anselin L, 1995. Local indicators of spatial association-LISA. Geogr Anal 27:93-115.

Bray F, Ferlay J, Soerjomataram I, Siegel R, Torre L, Jemal A, 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 394-424.

Chen W, Sun K, Zheng R, Zeng H, Zhang S, Xia C, Yang Z, Li H, Zou X, He J, 2018. Cancer incidence and mortality in China, 2014. Chin J Cancer Res 30:1-12.

Chen W, Zheng R, 2015. Incidence, mortality and survival analysis of breast cancer in China. Chin J Clin Oncol 668-74.

Chu J, Zhou C, Guo X, Sun J, Xue F, Zhang J, Lu Z, Fu Z, Xu A, 2017. Female breast cancer mortality clusters in Shandong Province, China: a spatial analysis. Sci Rep-Uk 7:1-08.

Chung S, Imberger J, Hipsey MR, Lee HS, 2014. The influence of physical and physiological processes on the spatial heterogeneity of a Microcystis bloom in a stratified reservoir. Ecol Model 289:133-49.

Debarsy N, Dossougoin C, Ertur C, Gnabo J, 2018. Measuring sovereign risk spillovers and assessing the role of transmission channels: a spatial econometrics approach. J Econ Dyn Control 87:21-45.

Elhorst JP, 2010. Applied spatial econometrics: raising the bar. Spat Econ Anal 5:9-28.

Fan C, Cai T, Gai Z, Wu Y, 2020. The relationship between the migrant population's migration network and the risk of COVID-19 transmission in China - Empirical analysis and prediction in prefecture-level cities. Int J Env Res Pub He 17:2630.

Fan H, Gao B, Xu R, Wang J, 2017. Optimization of Shanghai marine environment monitoring sites by integrating spatial correlation and stratified heterogeneity. Acta Oceanol Sin 36:111-21.

Fan L, Strasser-Weippl K, Li J, St Louis J, Finkelstein DM, Yu K, Chen W, Shao Z, Goss PE, 2014. Breast cancer in China. Lancet Oncol 15:e279-89.

Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, Bray F, 2021. Cancer statistics for the year 2020: An overview. Int J Cancer 149:778-89.

Fu J, Jiang D, Huang Y, 2014. China km grid population distribution data set. Glob Change Sci Res Data 1:1-04.

Ge E, Zhang X, Wang X, Wei X, 2016. Spatial and temporal analysis of tuberculosis in Zhejiang Province, China, 2009-2012. Infect Dis Poverty 5:11-20.

Ghoncheh M, Pournamdar Z, Salehiniya H, 2016. Incidence and mortality and epidemiology of breast cancer in the world. Asian Pac J Cancer Prev 17:43-46.

Graif C, Sampson RJ, 2009. Spatial heterogeneity in the effects of immigration and diversity on neighborhood homicide rates. Homicide Stud 13:242-60.

Halunga AG, Orme CD, Yamagata T, 2017. A heteroskedasticity robust Breusch - Pagan test for Contemporaneous correlation in dynamic panel data models. J Econometrics 198: 209-30.

Huang Y, Jiang D, Fu J, 2014. China km grid GDP distribution data set. Glob Change Sci Res Data 1:5-08.

Igene H, 2008. Global health inequalities and breast cancer: an impending public health problem for developing countries. Breast J 14:428-34.

Ji P, Gong Y, Jiang CC, Hu X, Di GH, Shao ZM, 2020. Association between socioeconomic factors at diagnosis and survival in breast cancer: A population□based study. Cancer Med-US 9:1922-36.

Kloog I, Haim A, Portnov BA, 2009. Using kernel density function as an urban analysis tool: Investigating the association between nightlight exposure and the incidence of breast cancer in Haifa, Israel. Comput Enviro Urban Syst 33:55-63.

Lantz PM, Mujahid M, Schwartz K, Janz NK, Fagerlin A, Salem B, Liu L, Deapen D, Katz SJ, 2006. The influence of race, ethnicity, and individual socioeconomic factors on breast cancer stage at diagnosis. Am J Public Health 96:2173-78.

Lei S, Zheng R, Zhang S, Wang S, Chen R, Sun K, Zeng H, Zhou J, Wei W, 2021. Global patterns of breast cancer incidence and mortality: a population-based cancer registry data analysis from 2000 to 2020. Cancer Commun 41:1183-94.

Li J, Zhang B, Fan J, Pang Y, Zhang P, Wang S, Zheng S, Zhang B, Yang H, Xie X, 2011. A nation-wide multicenter 10-year (1999-2008) retrospective clinical epidemiological study of female breast cancer in China. BMC Cancer 11:364.

Liao Y, Li D, Zhang N, Xia C, Zheng R, Zeng H, Zhang S, Wang J, Chen W, 2019. Application of sandwich spatial estimation method in cancer mapping: A case study for breast cancer mortality in the Chinese mainland, 2005. Stat Methods Med Res 28:3609-26.

Liao Y, Wang J, Wu J, Driskell L, Wang W, Zhang T, Xue G, Zheng X, 2010. Spatial analysis of neural tube defects in a rural coal mining area. Int J Environ Health Res 20:439-50.

Liu W, Song Z, Liu Z, 2016. Progress of economic geography in China's mainland since 2000. J Geogr Sci 26:1019-40.

Macquillan EL, Curtis AB, Baker KM, Paul R, Back YO, 2017. Using GIS mapping to target public health interventions: examining birth outcomes across GIS techniques. J Commun Health 42:633-38.

Nelson JK, Brewer CA, 2017. Evaluating data stability in aggregation structures across spatial scales: revisiting the modifiable areal unit problem. Cartogr Geogr Inf Sci 44:35-50.

Nordstokke DW, Zumbo BD, 2010. A new nonparametric Levene test for equal variances. Psicológica 31:401-30.

Qi F, Du F, 2013. Tracking and visualization of space-time activities for a micro-scale flu transmission study. Int J Health Geogr 12:6.

Qi Y, Yang Y, Jin F, 2013. China's economic development stage and its spatio-temporal evolution: A prefectural-level analysis. J Geogr Sci 23:297-314.

Ramadan AAB, Jacksonthompson J, Boren SA, 2017. Geographic information systems: usability, perception, and preferences of public health professionals. Online J Public Health Inform 9:1-16.

Ren J, Gong Y, Ling H, Hu X, Shao Z, 2019. Racial/ethnic differences in the outcomes of patients with metastatic breast cancer: contributions of demographic, socioeconomic, tumor and metastatic characteristics. Breast Cancer Res Treat 173:225-37.

Salinas GD, Whitworth L, Merwin P, Emarine J, 2018. Pancreatic cancer management and treatment landscape awareness of gastroenterologists: results from US physician surveys conducted in 2013 and 2015. J Gastrointest Cancer 49:41-9.

Seidman H, Stellman SD, Mushinski MH, 1982. A different perspective on breast cancer risk factors: some implications of the non-attributable risk. Cancer J Clin 32:301-13.

Shrestha A, Luo W, 2018. Assessment of groundwater nitrate pollution potential in Central Valley Aquifer using geodetector-based frequency ratio (GFR) and optimized-DRASTIC methods. ISPRIS Int J Geoinf 7:211.

Siegel RL, Miller KD, Jemal A, 2015. Cancer statistics, 2015. Cancer J Clin 65:5-29.

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F, 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Cancer J Clin 71:209-49.

Sun X, Zhao D, Liu Y, Liu Y, Yuan Z, Wang J, Xue F, 2017. The long-term spatial-temporal trends and burden of esophageal cancer in one high-risk area: a population-registered study in Feicheng, China. PLos One 12:e173211.

Tan J, Wang X, Pan J, 2021. The effect of population distribution measures on evaluating spatial accessibility of primary healthcare institutions: a case study from China. Geospat Health 16:32-45.

Wang J, Haining R, Liu T, Li L, Jiang C, 2013. Sandwich estimation for multi-unit reporting on a stratified heterogeneous surface. Environ Plann 45:2515-34.

Wang J, McMichael AJ, Meng B, Becker NG, Han W, Glass K, Wu J, Liu X, Liu J, Li X, 2006. Spatial dynamics of an epidemic of severe acute respiratory syndrome in an urban area. Bull World Health Organ 84:965-68.

Wang J, Stein A, Gao B, Ge Y, 2012. A review of spatial sampling. Spat Stat-Neth 2:1-14.

Wang J, Zhang T, Fu B, 2016. A measure of spatial stratified heterogeneity. Ecol Indic 67:250-6.

Wang JF, Li XH, Christakos G, Liao YL, Zhang T, Gu X, Zheng XY, 2010. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. Int J Geogr Inf Sci 24:107-27.

Wang Q, Li J, Zheng S, Li J, Pang Y, Huang R, Zhang B, Zhang B, Yang H, Xie X, 2012. Breast cancer stage at diagnosis and

area-based socioeconomic status: a multicenter 10-year retrospective clinical epidemiological study in China. BMC Cancer 12:122.
Xia C, Kahn C, Wang J, Liao Y, Chen W, Yu XQ, 2016. Temporal trends in geographical variation in breast cancer mortality in China, 1973-2005: an analysis of nationwide surveys on cause of death. Int J Env Res Pub Health 13:963.
Xu X, Zhao Y, Xia S, Zhang X, 2019. Investigation of multi-scale spatio-temporal pattern of oldest-old clusters in China on the basis of spatial scan statistics. PLoS One 14:e219695.
Xu X, Zhao Y, Zhang X, Xia S, 2018. Identifying the impacts of social, economic, and environmental factors on population aging in the Yangtze River Delta using the geographical detector technique. Sustainability-Basel 10:1528.
Zhou W, Zhang Z, Bi Y, Wang L, Liu X, Yu C, 2018. Joinpoint regression analysis for the mortality trend for breast cancer in Chinese female from 1987 to 2014. J Cent South Univ 43:210-5.