# Spatial Dataset of Breast Cancer Incidence Rate in China (2014–2016)

Wang, R. H.[1,2]   Wang, P. H.[1,2]   Xu, C. D.[1,2]   Wang, W.[1,2], Wang, Z. B.[1,2*]

1. Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract:** The authors used data from the China Cancer Registry Annual Report from 2017–2019, organizing the breast cancer incidence rates recorded in various cancer registries nationwide for the years 2014–2016. We vectorized and spatially visualized the breast cancer incidence rates at the county level in China using ArcGIS software. Descriptive and spatial statistical analyses were employed as study methods to explore the regional differences, spatial distribution, and trends of the changes in the breast cancer incidence rates in China. A dataset on the spatial distribution of breast cancer incidence at the county level in China for 2014–2016 was constructed, which contains the following: (1) county-level breast cancer incidence data; (2) grouped statistical data of county-level breast cancer incidence; (3) statistical data of breast cancer incidence in eastern, central, and western regions. The dataset is archived in .shp and .xlsx formats, consisting of 25 data files, with a total size of 21.5 MB.

**Keywords:** China; breast cancer; incidence; spatial distribution

## 1 Introduction

Breast cancer is one of the most common cancers worldwide, with the highest incidence rates observed in developed regions such as North America, Europe, and Australia[1]. In China, the incidence of breast cancer is relatively low compared with the global levels; however, over the past 30 years, this incidence has increased by 20%–30%, for an annual growth rate of approximately 3%–5%, surpassing the global average growth rate of 1.5%[2]. The national cancer statistics released by the National Cancer Center (NCC) of China

**Data Citation:** [1] Wang, R. H., Wang, P. H., Xu, C. D., *et al*. Spatial dataset of breast cancer incidence rate in China (2014–2016) [J]. *Journal of Global Change Data & Discovery*, 2024, 8(2): 147–154. https://doi.org/10.3974/geodp.2024.02.04. https://cstr.escience.org.cn/CSTR:20146.14.2024.02.04.

[2] Wang, R. H., Wang, P. H., Xu, C. D., *et al*. Spatial dataset of breast cancer incidence rate in China (2014–2016) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2024. https://doi.org/10.3974/geodb.2024.06.06.V1. https://cstr.escience.org.cn/CSTR:20146.11.2024.06.06.V1.

indicate that breast cancer has the highest incidence among the cancers affecting Chinese women[3]. The burden due to breast cancer in China may become more severe in the future with rapid economic and social development, population growth, the aging trend, and major prevalent risk factors[4]. Therefore, research on breast cancer in China is urgently needed to address and control the personal and socioeconomic burdens associated with breast cancer risk.

China is a vast territory with wide differences in its socioeconomic conditions and natural environments across various regions. Treating China as a single entity overlooks the regional heterogeneity in how risk factors affect disease incidence, potentially leading to research conclusions that do not align with the actual situation[5]. Most recent studies on the spatial distribution of and factors influencing breast cancer in China have been limited to small areas such as individual cities or provinces[6–9]. Nationwide studies have been relatively scarce and have mainly been conducted at the provincial level[10–12]. This limitation regarding spatial scale may have resulted in a lack of precision in the results, so the true breast cancer situation in China may not have been fully reflected.

In this study, we used counties as the unit, organizing breast cancer incidence data from various cancer registries in the China Cancer Registry Annual Report from 2017 to 2019[13–15]. These data were converted into vector data using ArcGIS 10.8 software to create a dataset. With this dataset, we explored the regional differences in, the spatial distribution of, and the trends in the changes in in breast cancer incidence in China using statistical charts, vector data visualization, and spatial autocorrelation analysis. This study provides a scientific basis for formulating targeted breast cancer prevention and control strategies.

## 2 Metadata of the Dataset

The metadata from the Dataset of spatial distribution of breast cancer incidence in China (2014–2016) [16] are summarized in Table 1, including the dataset name, authors, geographical region, year of the dataset, data files, data publication and sharing service platform, and data sharing policy, etc.

## 3 Methods

### 3.1 Data Sources

The breast cancer data used in this study were obtained from the China Cancer Registry Annual Report published by the National Cancer Center (NCC)[13–15]. The data cover the period from 2014 to 2016, and include 339, 388, and 487 cancer registry areas, respectively, across all 31 provinces in mainland China. The populations included in this study were 288 million, 321 million, and 382 million, accounting for 21.07%, 23.35%, and 27.6% of China's end-of-year population in 2014–2016, respectively. The women accounted for 142 million, 158 million, and 187 million of these values, respectively. Among these, 59, 806, 67, 328, and 79, 450 were new breast cancer cases in women, respectively.

The annual report includes two types of breast cancer incidence rate statistics: (1) The crude incidence rate is the basic indicator of the incidence in the population. This refers to the number of new cancer cases registered per 100,000 people in a specific year in a specific area and reflects the population incidence level. (2) The age-standardized rate (ASR) is adjusted for the age structure of a given standard population to eliminate the impact of age structure on the incidence level. Because the crude incidence rate is strongly affected by the age composition of the population, we selected the age-standardized incidence rate of breast cancer (units: per 100,000) from the annual report as the statistical indicator.

**Table 1**  Metadata summary of the dataset

| Items | Description |
|---|---|
| Dataset full name | Spatial dataset of breast cancer incidence rate in China (2014–2016) |
| Dataset short name | BreastCancerIR2014-2016 |
| Authors | Wang, R. H., Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, wangruohan2446@igsnrr.ac.cn |
| | Wang, P. H., Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, wph1996@126.com |
| | Xu, C. D., Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, xucd@lreis.ac.cn |
| | Wang, W., Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, wang_wei@lreis.ac.cn |
| | Wang, Z. B., Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, wangzb@igsnrr.ac.cn |
| Geographical region | China |
| Years | 2014–2016 |
| Data format | .shp, .xlsx |
| Data size | 21.5 MB |
| Data files | Spatial distribution vector data and tabulated statistical data on breast cancer incidence in China (2014–2016) |
| Foundations | National Natural Science Foundation of China (42130713); Ministry of Science and Technology of P. R. China (2019QZKK1005) |
| Data publisher | Global Change Research Data Publishing & Repository, http://www.geodoi.ac.cn |
| Address | No. 11A, Datun Road, Chaoyang District, Beijing 100101, China |
| Data sharing policy | (1) *Data* are openly available and can be free downloaded via the Internet; (2) End users are encouraged to use *Data* subject to citation; (3) Users, who are by definition also value-added service providers, are welcome to redistribute *Data* subject to written permission from the GCdataPR Editorial Office and the issuance of a *Data* redistribution license; and (4) If *Data* are used to compile new datasets, the 'ten per cent principal' should be followed such that *Data* records utilized should not surpass 10% of the new dataset contents, while sources should be clearly noted in suitable places in the new dataset[17] |
| Communication and searchable system | DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS, GEOSS, PubScholar, CKRSC |

## 3.2 Classification

We descriptively and statistically analyzed the breast cancer incidence rates according to the regional classification method in the 2019 China Cancer Registry Annual Report[15]. The regional classification standards were as follows:

(1) Eastern region: Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, and Hainan (11 provinces/municipalities).

(2) Central region: Shanxi, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei, and Hunan (eight provinces).

(3) Western region: Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, and Xinjiang (12 provinces/municipalities/autonomous regions).

## 3.3 Data Processing

The breast cancer incidence data for the tumor registry areas in 2014, 2015, and 2016 were extracted from the China Cancer Registry Annual Report of 2017, 2018, and 2019, respectively. These data were then summarized in .xlsx file as source data for the descriptive statistics; the data were then spatially visualized using ArcGIS10.8 software, serving as the source data for spatial statistical analysis. Considering the matching between the statistical areas of each cancer registry and the attribute table of China's county-level vector data, the county-level vector data for 2014–2016 used in this study included 339, 386, and 485 districts/counties, respectively.

### 3.4 Spatial Statistical Analysis

Geographical phenomena commonly exhibit spatial dependence: geographically close areas or observation points often show a degree of similarity or correlation. The Moran's I statistic, which is widely used to test for global spatial autocorrelation, was thus applied to quantify this spatial dependence. The Moran's I statistic measures the average similarity between each spatial observation and its nearby observations, with values ranging from –1 to 1[18]. A Moran's I of greater than 0 denotes a positive spatial autocorrelation, indicating that observations in neighboring areas tend to be similar; the larger the value, the stronger the correlation. A Moran's I of less than 0 denotes a negative spatial autocorrelation, indicating that observations in neighboring areas tend to show opposite trends. When Moran's I equals 0, the spatial distribution of the observations is essentially random, with no spatial autocorrelation. Calculating Moran's I helps with understanding the spatial characteristics and distribution patterns of geographical phenomena, providing a scientific basis for further spatial analysis and decision-making.

Moran's I formula is as follows：

$$I = \frac{n\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}w_{ij}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2} \tag{3}$$

where $x$ represents the observation variable, $x_i$ and $x_j$ are the attribute values of the $i$-th and $j$-th spatial units respectively (for example, the breast cancer incidence rate of the $i$-th county-level city and the breast cancer incidence rate of the $j$-th county-level city), $w_{ij}$ is the spatial weight between variables $i$ and $j$, and $n$ is the total number of variables.

## 4 Data Results

### 4.1 Descriptive Statistical Analysis

Cancer registration involves long-term, continuous, and dynamic systematic monitoring of cancer prevalence, trends, and influencing factors. Cancer registration is thus a crucial foundational task for developing cancer prevention and control strategies, conducting comprehensive research, and evaluating the effectiveness of cancer prevention and control[15]. From 2014 to 2016, the National Cancer Center included 339, 388, and 487 cancer registries, showing the expansion of its coverage. This expansion reflects the country's emphasis on cancer prevention and control and the health of its people. The number and distribution of the districts and counties used in this study are shown in Table 2.

**Table 2**　Number of cancer registries from 2014 to 2016

|           | 2014 | 2015 | 2016 |
|-----------|------|------|------|
| Number    | 339  | 386  | 485  |
| Increment |      | 47   | 99   |

First, a descriptive statistical analysis was performed on county-level breast cancer incidence rates in China. The resulting histogram shows that between 2014 and 2016, most counties in China had an age-standardized incidence rate (ASR) of breast cancer of below 35, with the highest number of counties having an ASR of between 15 and 30, indicating that the overall breast cancer incidence rate in China was relatively stable (Figure 1). However, the box plot reveals some variation in breast cancer incidence rates among the counties in

eastern, central, and western China (Figure 2).
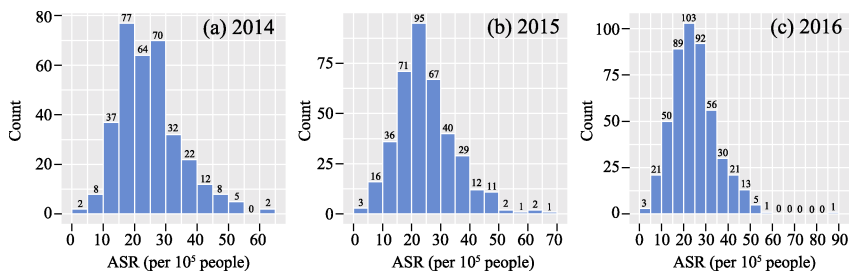


**Figure 1**    Histogram of breast cancer incidence rates in China: (a) 2014; (b) 2015; (c) 2016
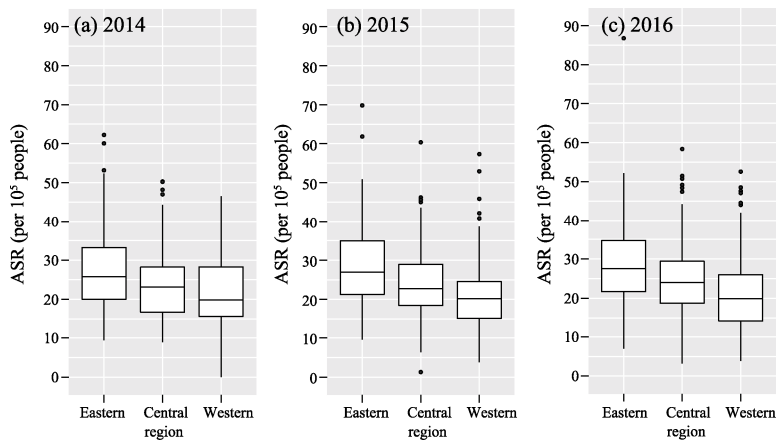


**Figure 2**    Box plot of breast cancer incidence rates in Eastern, Central, and Western China: (a) 2014; (b) 2015; (c) 2016

Table 3 shows that breast cancer incidence rates in the eastern region were generally higher than those in the central and western regions. The median incidence rates in the eastern region increased annually, with the largest variance, indicating substantial spatial heterogeneity in the incidence rates across the counties. Although the median incidence rate in the central region varied little, the variance increased annually, suggesting a possible trend toward more pronounced spatial differentiation in this region. The median incidence rate in the western region was considerably lower than those in the eastern and central regions, and the variance was relatively stable, indicating that the incidence rates in this area were relatively stable.

**Table 3**    Incidence rate of breast cancer in Eastern, Central and Western China, 2014–2016.

| Region | Year | Max | Min | Median | Var |
|---|---|---|---|---|---|
| Eastern | 2014 | 62.23 | 9.37 | 25.79 | 100.57 |
|  | 2015 | 69.81 | 9.57 | 26.94 | 101.41 |
|  | 2016 | 86.75 | 6.96 | 27.54 | 100.67 |
| Central | 2014 | 50.31 | 8.91 | 23.12 | 82.56 |
|  | 2015 | 60.36 | 1.30 | 22.75 | 86.70 |
|  | 2016 | 58.35 | 3.19 | 24.06 | 91.90 |
| Western | 2014 | 46.51 | 0.00 | 19.84 | 89.29 |
|  | 2015 | 57.33 | 3.79 | 20.20 | 89.28 |
|  | 2016 | 52.56 | 3.88 | 19.89 | 89.60 |

## 4.2 Spatial Statistical Analysis

From 2014 to 2015, Liaoning, Shandong, Henan, Shanghai, and Shenzhen had higher breast cancer incidence rates. In 2016, the high-incidence areas expanded to include Beijing, Hainan, Inner Mongolia, and some regions in western China. Regions with lower incidence rates were primarily located in the central and western parts of China and in Jiangsu and Fujian in the eastern region (Figure 3). The results of the spatial autocorrelation analysis indicated that from 2014 to 2016, the *p*-value for Moran's I of breast cancer incidence rates in China was consistently less than 0.000,1, demonstrating strong spatial clustering of the breast cancer incidence rates (Table 4).
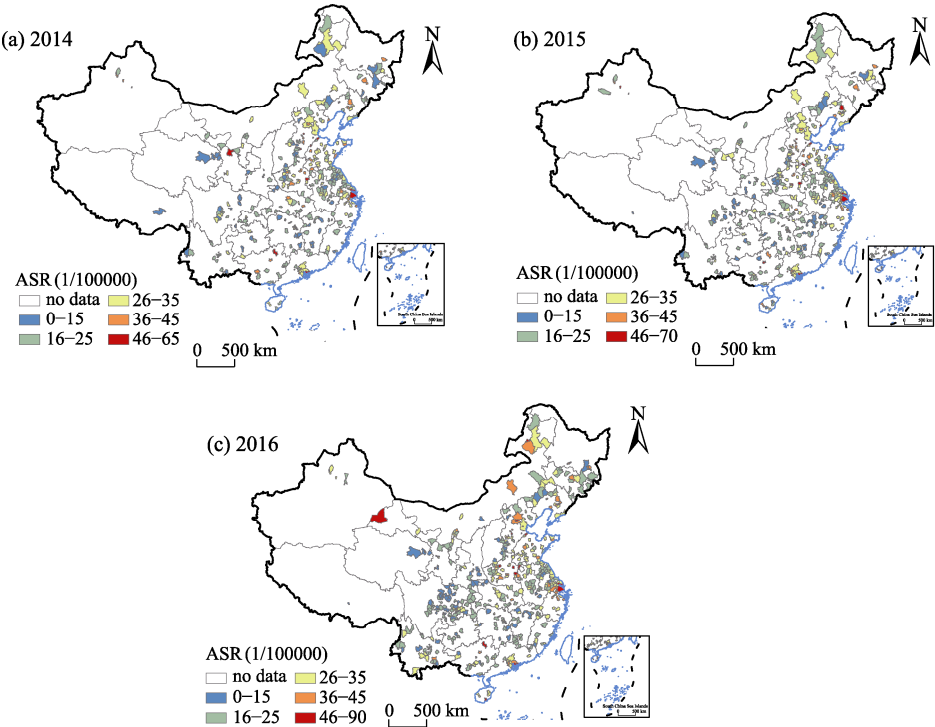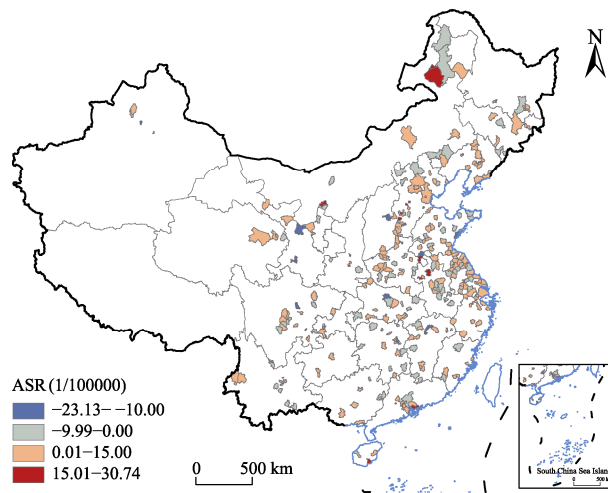


**Figure 3**   Maps of breast cancer incidence rates in China (2014–2016)

**Table 4**   Global spatial autocorrelation analysis of breast cancer incidence rates in China (2014–2016)

| Year | Moran's I | z-score | *p*-value | Spatial Pattern |
| --- | --- | --- | --- | --- |
| 2014 | 0.17 | 7.12 | <0.000,1 | clustered |
| 2015 | 0.12 | 13.15 | <0.000,1 | clustered |
| 2016 | 0.16 | 20.62 | <0.000,1 | clustered |

Figure 4 shows that breast cancer incidence rates were higher in most regions in China in 2016 than in 2014. Notably, the incidence rates increased in areas such as Beijing–Tianjin, Liaoning, southern Hebei, the Yangtze River Delta, and the Pearl River Delta, with some increases observed in parts of southwest and northwest China. Conversely, the incidence rates decreased in certain counties in northern Hebei, eastern Shandong, and central Jiangsu in the eastern region; in central–southern Anhui, Hunan, and Hubei in the central region; and in some counties in Sichuan, Yunnan, and Gansu in the western region.

**Figure 4**　Analysis map of changes in breast cancer incidence rates in China (2014−2016)

## 5 Discussion and Conclusions

Based on data from the China Cancer Registry Annual Report from 2017 to 2019, we constructed a dataset that provides vector data on county-level breast cancer incidence rates in China for the years 2014–2016, enhancing the spatial precision of the previously published rates. Using descriptive and spatial statistical methods, we thoroughly investigated the regional differences and spatial trends in breast cancer incidence rates across China.

The results indicated that the breast cancer incidence rates in China exhibited considerable regional differences, with rates decreasing from the east to central and then to western regions, with notable spatial clustering. From 2014 to 2016, regions with higher breast cancer incidence rates expanded from Liaoning, Shandong, Henan, Shanghai, and Shenzhen to Beijing, Hainan, Inner Mongolia, and some areas in western China. Conversely, regions with lower incidence rates remained consistently distributed in the central and western regions, as well as in Jiangsu and Fujian in the eastern region. From 2014 to 2016, breast cancer incidence rates increased in most regions in China, particularly in areas such as Beijing–Tianjin, Liaoning, southern Hebei, the Yangtze River Delta, and the Pearl River Delta. Some counties in southwest and northwest China also experienced rising incidence rates. However, counties with decreasing incidence rates were primarily located in northern Hebei, eastern Shandong, and central Jiangsu in the eastern region; central–southern Anhui, Hunan, and Hubei in the central region; and Sichuan, Yunnan, Gansu, and Ningxia in the western region. In some high-incidence areas, such as Henan and Shandong, breast cancer incidence rates decreased over the study period. Conversely, regions with initially low incidence rates, such as southern Hebei, Jiangsu, northern Anhui, and Ningxia, showed an upward trend in their rates. This indicates that while monitoring, prevention, and treatment efforts should continue in high-incidence areas, increased attention must be paid to the rising incidence rates in lower-incidence areas. This focus will help to promote the rational allocation of healthcare resources, enhance public health awareness, and gradually control and reduce the national and individual burdens due to breast cancer.

The dataset covers approximately 17% of all counties in China; therefore, the true breast cancer incidence rates across the entire country may not be fully represented. However, the results can be considered reasonably representative given the overall volume of data. Future studies should build on these findings to further analyze the spatial variation in breast cancer incidence rates and the underlying mechanisms for these variations. This information will

help with devising more effective and rational policies and methods for controlling and mitigating the various negative effects of breast cancer.

### *Author Contributions*
Wang, R. H. developed the overall design for the construction of the dataset; Wang, R. H. collected and processed the data and wrote the paper; Wang, R. H., Wang, P. H., Xu C. D., Wang W., and Wang Z. B. validated the data.

### *Conflicts of Interest*
The authors declare no conflicts of interest.

## References

[1]    Sung, H., Ferlay, J., Siegel, R. L., *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. *CA-A Cancer Journal for Clinicians*, 2021, 71(3): 209−249.

[2]    Li, T., Mello-Thoms, C., Brennan, P. C. Descriptive epidemiology of breast cancer in China: incidence, mortality, survival and prevalence [J]. *Breast Cancer Research and Treatment*, 2016, 159(3): 395−406.

[3]    Zheng, R., Zhang, S., Zeng, H., *et al*. Cancer incidence and mortality in China, 2016 [J]. *Journal of the National Cancer Center*, 2022, 2(1): 1−9.

[4]    Lei, S., Zheng, R., Zhang, S., *et al*. Breast cancer incidence and mortality in women in China: temporal trends and projections to 2030 [J]. *Cancer Biology & Medicine*, 2021, 18(3): 900−909.

[5]    Wang, J. F., Zhang, T. L., Fu, B. J. A measure of spatial stratified heterogeneity [J]. *Ecological Indicators*, 2016, 67: 250−256.

[6]    Fei, X., Lou, Z., Christakos, G., *et al*. A geographic analysis about the spatiotemporal pattern of breast cancer in Hangzhou from 2008 to 2012 [J]. *Plos One*, 2016, 11(1): 1−13.

[7]    Song, M. J., Huang, X. X., Wei, X. Q., *et al*. Spatial patterns and the associated factors for breast cancer hospitalization in the rural population of Fujian Province, China [J]. *BMC Womens Health*, 2023, 23(1): 247−255.

[8]    Huo, Q., Zhang, N., Wang, X., *et al*. Effects of ambient particulate matter on human breast cancer: is xenogenesis responsible? [J]. *Plos One*, 2013, 8(10): e76609−e76615.

[9]    Yu, Q., Zhang, L., Hou, K., *et al*. Relationship between air pollutant exposure and gynecologic cancer risk [J]. *International Journal of Environmental Research and Public Health*, 2021, 18(10): 5353−5366.

[10]   He, R., Zhu, B., Liu, J., *et al*. Women's cancers in China: a spatio-temporal epidemiology analysis [J]. *BMC Womens Health*, 2021, 21(1): 116−129.

[11]   Xia, C., Kahn, C., Wang, J., *et al*. Temporal trends in geographical variation in breast cancer mortality in China, 1973−2005: an analysis of nationwide surveys on cause of death [J]. *International Journal of Environmental Research and Public Health*, 2016, 13(10): 963−978.

[12]   Hu, M. Y., Jiang, C., Meng, R. T., *et al*. Effect of air pollution on the prevalence of breast and cervical cancer in China: a panel data regression analysis [J]. *Environmental Science and Pollution Research*, 2023, 30(34): 82031−82044.

[13]   National Cancer Center. 2017 China Cancer Registry Annual Report [M]. Beijing: People's Health Publishing House, 2018.

[14]   National Cancer Center. 2018 China Cancer Registry Annual Report [M]. Beijing: People's Health Publishing House, 2019.

[15]   National Cancer Center. 2019 China Cancer Registry Annual Report [M]. Beijing: People's Health Publishing House, 2021.

[16]   Wang, R. H., Wang, P. H., Xu C. D., *et al*. Spatial dataset of breast cancer incidence rate in China (2014−2016) [J/DB/OL]. *Digital Journal of Global Change Data Repository*, 2024. https://doi.org/10.3974/geodb.2024.06.06.V1. https://cstr.escience.org.cn/CSTR:20146.11.2024.06.06.V1.

[17]   GCdataPR Editorial Office. GCdataPR data sharing policy [OL]. https://doi.org/10.3974/dp.policy.2014.05 (Updated 2017).

[18]   Moran, P. A. P. Notes on continuous stochastic phenomena [J]. *Biometrika*, 1950, 37(1/2): 17−23.