

## RESEARCH ARTICLE

# Detecting spatio-temporal hotspots of scarlet fever in Taiwan with spatio-temporal $G_i^*$ statistic

Jia-Hong Tang<sup>1</sup>, Tzu-Jung Tseng<sup>2</sup>, Ta-Chien Chan<sup>1,2\*</sup>

**1** Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, **2** Research Center for Humanities and Social Sciences, Academia Sinica, Taipei, Taiwan

\* [dachianpig@gmail.com](mailto:dachianpig@gmail.com)



## Abstract

A resurgence of scarlet fever has caused many pediatric infections in East Asia and the United Kingdom. Although scarlet fever in Taiwan has not been a notifiable infectious disease since 2007, the comprehensive national health insurance data can still track its trend. Here, we used data from the open data portal of the Taiwan Centers for Disease Control. The scarlet fever trend was measured by outpatient and hospitalization rates from 2009 to 2017. In order to elucidate the spatio-temporal hotspots, we developed a new method named the spatio-temporal  $G_i^*$  statistic, and applied Joinpoint regression to compute the annual percentage change (APC). The overall APCs in outpatient and hospitalization were 15.1% (95% CI: 10.3%-20.2%) and 7.7% (95%CI: 4.5% -10.9%). The major two infected groups were children aged 5–9 (outpatient: 0.138 scarlet fever diagnoses per 1,000 visits; inpatient: 2.579 per 1,000 visits) and aged 3–4 (outpatient: 0.084 per 1,000 visits; inpatient: 1.469 per 1,000 visits). We found the counties in eastern Taiwan and offshore counties had the most hotspots in the outpatient setting. In terms of hospitalization, the hotspots mostly occurred in offshore counties close to China. With the help of the spatio-temporal statistic, health workers can set up enhanced laboratory surveillance in those hotspots.

## OPEN ACCESS

**Citation:** Tang J-H, Tseng T-J, Chan T-C (2019) Detecting spatio-temporal hotspots of scarlet fever in Taiwan with spatio-temporal  $G_i^*$  statistic. PLoS ONE 14(4): e0215434. <https://doi.org/10.1371/journal.pone.0215434>

**Editor:** Ram K. Raghavan, Kansas State University, UNITED STATES

**Received:** July 28, 2018

**Accepted:** April 2, 2019

**Published:** April 16, 2019

**Copyright:** © 2019 Tang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data we used in this study were all third-party data available from the Taiwan CDC open data portal (<https://data.cdc.gov.tw/en/dataset/hi-outpatient-emergency-visit-scarlet-fever>). We confirm that anyone can access and download without any special access privileges.

**Funding:** TCC received the grant. This research was supported by a grant titled "Multidisciplinary Health Cloud Research Program: Technology Development and Application of Big Health Data" from Academia Sinica (<https://www.sinica.edu.tw/>)

## Introduction

*Streptococcus pyogenes* causes a variety of human diseases, including relatively mild skin infections as well as severe invasive diseases [1]. Among the diseases caused by this pathogen, scarlet fever, characterized by a sore throat, fever, headaches, swollen lymph nodes, and a characteristic rash, is predominantly an infectious disease of childhood, though it can also occur in older children and adults [2]. With improved nutrition and widespread use of antibiotics, scarlet fever is now a common, mild contagious disease. However, it is still a notifiable disease in many countries and regions. During the last decade, sporadic outbreaks and reemerging epidemics have been recorded worldwide, including in Vietnam [3], the Republic of Korea [4], China [5–7], Hong Kong [8, 9], Australia [10], Poland [11], and the United Kingdom [12, 13]. In Taiwan, scarlet fever was removed in 2007 from the list of notifiable diseases because of improved medical care capacities [14]. However, through using national health insurance data, the Taiwan Centers for Disease Control (Taiwan CDC) can still monitor the morbidity and hospitalization trends of scarlet fever.

en). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

Understanding the distribution of a disease in time and space is a foundation of public health. One of the most relevant analyses is related to analyzing the aggregation of disease cases in space, in time, or in both space and time [15–17]. In epidemiological studies, it is important to evaluate whether variation in the mortality and morbidity of a disease is randomly distributed or tends to occur as clusters over time and space in order to find the causative mechanism of the disease [18, 19]. Quantifying spatio-temporal patterns is important to identify the hotspots for public health intervention. Spatio-temporal statistics and tests are useful for adding precision to qualitative descriptions, facilitating the comparison of distributions, and drawing attention to some characteristics that might not be easily identified upon visual inspection.

The mapping of disease incidence and prevalence has long been a part of public health, epidemiology, and the study of disease in human populations [20]. Increasingly for epidemiological study, researchers have been looking for novel data visualization methods to aid in exploring spatial and temporal patterns. The ring map facilitates the visual assessment of multivariate spatial data by depicting individual datasets as separate rings of information surrounding a base map of a particular geographic region of interest [21, 22]. In this way, a ring map effectively summarizes multiple layers of data, presenting an array of regional attributes in a single spatially referenced graphic; ring maps have been used largely for representation of temporal data.

Local spatial statistics, such as local Moran's  $I$  [23], the Getis-Ord  $G_i^*$  statistic [24, 25] and the score statistic [26], may assist with the identification of disease clusters. Local spatial statistics identify both those clusters with values higher in magnitude than are expected to be found by random chance, and the statistically significant patterns of high risk (hotspots) or low risk (cold spots) frequency locations. Such statistics require that a neighborhood can be defined around the location of interest. This is done by specifying weights for surrounding regions, and is tantamount to specification of the scale at which the local dependence or clustering is tested.

The Getis-Ord  $G_i^*$  hotspot cluster statistic is one of the popular approaches used for local spatial analysis. The  $G_i^*$  statistic measures the degree of spatial clustering of a local sample and how different it is from the expected value, which is the mean of the whole data set. Study of the annual expansion of disease clusters can use the annual maps of hotspots and follow their expansion. Yet, since the  $G_i^*$  statistic is a measure relative to the overall mean in a particular year, and since the mean varies annually, the discovery of temporally related hotspots is constrained.

This paper introduces a spatio-temporal  $G_i^*$  statistic whereby the spatial or temporal features that the user wants to incorporate in the formulation are directly accounted for in the generation of neutral models. The objective of this paper is to investigate the spatio-temporal patterns for scarlet fever in Taiwan and use the proposed spatio-temporal  $G_i^*$  statistic for the detection of local clusters and anomalies in outpatients and hospital admissions. A ring map using ArcPy and Python language to develop the ArcToolbox (<http://www.esri.com/esri-news/arcuser/fall-2013/looking-at-temporal-changes>) was created to visually explore the spatio-temporal distribution of scarlet fever using the proposed  $G_i^*$  Z-Scores on one map and also to identify the spatio-temporal hotspots [27].

## Materials and methods

### Ethics

The data we used in this study were all from the Taiwan CDC open data portal (<https://data.cdc.gov.tw/en/>). The values are all the aggregated counts of scarlet fever cases and total visits

from outpatient and inpatient settings in each city or county. Consequently, approval from the institutional review board (IRB) was not required, and we also did not need to get consent from each patient.

### Data source

The Taiwan CDC open data portal, which can be accessed by the public, provides more than 250 datasets, including all notifiable disease cases and emergency department visits for selected syndromes, and information on quarantine practices, vaccines, nosocomial infections, and the like. Most of the information is updated automatically. More than 99% of the Taiwan population is covered by National Health Insurance (NHI). The Taiwan CDC receives aggregated numbers of outpatient, inpatient, and emergency room (ER) visits from NHI claims data via machine-to-machine interface, since 2008. In this study, yearly county-level data from scarlet fever cases in Taiwan were collected online during the period from 2009 to 2017, from Taiwan CDC's open data portal. The scarlet fever dataset comprises numbers of inpatients and outpatients by gender and age group to provide an understanding of epidemic situations and disease trends in different cities and counties.

### Statistical methods

**Temporal trend.** We used Joinpoint Regression Software version 4.5.0.1, developed by the National Cancer Institute, to examine annual percentage changes (APC) in overall and age-specific outpatient and hospitalization rates from 2009 to 2017 [7]. A permutation test procedure in Joinpoint regression was used to assess whether an annual percentage change was significantly different from zero and the  $p$ -value for a two-sided test [28]. In describing trends, an increase or decrease is proved when the slope of APC is statistically significant ( $p < 0.05$ ). When the slope of annual percentage change is significant ( $p < 0.05$ ), it indicates an increasing or decreasing trend. A stable trend means a non-significant annual percentage change within the period ( $p \geq 0.05$ ).

**Spatio-temporal trend.** In epidemiology, disease maps are often used to explore the hotspots of diseases [29, 30]. Disease incidence and prevalence contain both spatial and temporal attributes. One common approach to understanding spatial and temporal trends is to break the data up into a series of time snapshots. However, this usually leads to concerns of discontinuity.

The  $G_i^*$  statistic, introduced by Getis and Ord [24, 25], can be used as a measure of the degree of spatial clustering. Exploring the spatio-temporal pattern of the disease hotspots is limited with the  $G_i^*$  statistic since relative hotspots are accepted according to the distribution pattern of a specific period such as year, or month.

For the purpose of developing a modification of the  $G_i^*$  statistic which incorporates the spatio-temporal association and correlations, we consider both the spatial characteristic as well as the temporal characteristic in the formation of the spatio-temporal neighborhood. Common conceptualizations of spatial relationships include inverse distance, travel time, fixed distance,  $K$  nearest neighbors and contiguity. For the temporal characteristics, each observation at the current time and location is not only influenced by the previous time at the location, but also affected by the previous time of its spatial neighbors. A time lag is used to capture this feature. Then the spatial neighborhood and the time lag are combined to form the spatio-temporal neighborhood.

A modified  $G_i^*$  is developed in this study which identifies year-to-year hotspots which are relative to the past few years' observations. The spatio-temporal  $G_i^*$  statistic is proposed as

follows.

$$G_i^*(d, t) = \frac{\sum_{t=0}^l \sum_{j \in \partial i} w_{ij}(d, t) \cdot x_{jt} - \bar{x} \sum_{t=0}^l \sum_{j \in \partial i} w_{ij}(d, t)}{S \sqrt{\frac{\sum_{t=0}^l \sum_{j \in \partial i} w_{ij}^2(d, t) - \left( \sum_{t=0}^l \sum_{j \in \partial i} w_{ij}(d, t) \right)^2}{n-1}}} \tag{1}$$

where  $x_j$  is the specified attribute value of location  $j$ ,  $\partial i$  denotes the set of spatial neighbors of location  $i$  (including location  $i$ ),  $l$  is the designated time lag,  $n$  is the number of spatio-temporal neighbors of location  $i$  (including location  $i$ ),  $w_{ij}(d, t)$  is the spatio-temporal weight for spatio-temporal neighbor  $j$  from location  $i$ , and  $\bar{x}$  and  $S$  represents the mean and the standard deviation of the specified attribute of the set of spatio-temporal neighbors of location  $i$  (including location  $i$ ) and where

$$S = \sqrt{\frac{\sum_{t=0}^l \sum_{j \in \partial i} x_{jt}^2}{n} - \bar{x}^2} \tag{2}$$

When  $l = 0$ , this modified  $G_i^*$  statistic is the same as the original form. Our notion of a spatio-temporal neighborhood is distinct from the traditional notions, since we consider both a spatial characterization as well as a temporal characterization of neighborhoods. To identify and describe spatio-temporal association between location  $i$  and its spatio-temporal neighbors, the weight function is defined as follows:

$$w_{ij}(d, t) = \left( 1 - \frac{d_{ij}}{\sum_{j \in \partial i} d_{ij}} \right)^{t+1}, j \in \partial i, \tag{3}$$

where  $d_{ij}$  is the distance between locations  $i$  and  $j$ , and  $t$  represents time lags.

The proposed weight function is a kind of distance declining effect that decreases exponentially with time lags. The distance declining effect means that as distance increases from a particular location, the weight between spatio-temporal neighbors lessens. We use Z scores to measure the standard deviation; under the normal distribution, scores within one standard deviation from the average value will account for 68.27% of the data, those within two standard deviations will reach 95.45%, and those within three standard deviations will reach 99.73%.

Due to the spatio-temporal structure of the data, neighboring locations exist both in time and in space. In this study, the first-order Queens Case Contiguity method [31] was used to define neighborhood size in space. To define temporal neighbors, time-lagged correlations, which describe how similar the time series is with itself, were calculated and tested until the first statistically insignificant result appeared. The results of time lag selection in this study are shown in S1 and S2 Tables.

The R source code and the data we used are provided at the open repository, figshare.com (<https://figshare.com/s/cf8762f07e48ec79a4df>). The ring map and associated Taiwan base map created in this study are also at figshare.com (<https://figshare.com/s/8d2a3887482079d3e2bd>).

## Results

### Epidemiological trend of scarlet fever

In view of the Joinpoint regression results for the outpatient rate in Table 1, there was a significant upward trend of 15.1% (95% CI: 10.3%-20.2%) per year for the overall scarlet fever

**Table 1. Descriptive statistics on annual rate of scarlet fever in outpatient setting from 2009 to 2017 and annual percentage change (APC).**

Descriptive Statistics						
OPD Age group	N1	Rate Range	Mean (per 1,000)	Std. Deviation	N2	APC % (95%CI)
0–2	198	(0, 0.4)	0.0315	0.0446	9	12.4 (8.8, 16.2)
3–4	198	(0, 2.02)	0.0836	0.17327	9	17.9 (11.7, 24.3)
5–9	198	(0, 4.02)	0.138	0.30878	9	21.5 (15.2, 28.1)
10–14	198	(0, 0.23)	0.0262	0.03283	9	19.8 (13.3, 26.6)
15+	198	(0, 0.03)	0.0005	0.00237	9	3.1 (-1.5, 7.9)
Overall	198	(0, 0.18)	0.0120	0.01593	9	15.1 (10.3, 20.2)

Abbreviations: CI, confidence interval. Age is in years. N1 represents annual rate in 22 cities or counties for 9 years. N2 represents annual rate in the whole of Taiwan for 9 years.

<https://doi.org/10.1371/journal.pone.0215434.t001>

outpatient rate between 2009 and 2017. From the point of view of age-specification, APCs of the scarlet fever outpatient rate increased 12.4% (95% CI: 8.8%-16.2%), 17.9% (95% CI: 11.7%-24.3%), 21.5% (95% CI: 15.2%-28.1%), and 19.8% (95% CI: 13.3%-26.6%) per year from 2009 to 2017 for the age groups 0–2, 3–4, 5–9 and 10–14 years, respectively. For the age group 15+, APC of the scarlet fever outpatient rate was stable between 2009 and 2017. For the outpatient rates of the age group 5–9 years, there was higher deviation than for the other age groups.

In the light of the Joinpoint regression results for the inpatient rate in Table 2, there was a significant upward trend of 7.7% (95% CI: 4.5%-10.9%) per year for the overall scarlet fever inpatient rate between 2009 and 2017. For the age-specific perspective, APCs of the scarlet fever inpatient rate increased 12.8% (95% CI: 7.2%-18.6%), 16.0% (95% CI: 9.5%-22.9%) and 16.3% (95% CI: 9.4%-23.4%) per year from 2009 to 2017 for the age groups 3–4, 5–9 and 10–14, respectively. For the age groups 0–2 and 15+, APCs of the scarlet fever inpatient rate were stable between 2009 and 2017. For inpatient rates of the age groups 3–4 and 5–9, there were higher deviations than for the other groups. This shows that inpatient rates between 2009 and 2017 of the age groups 3–4 and 5–9 were more spread out and had larger ranges.

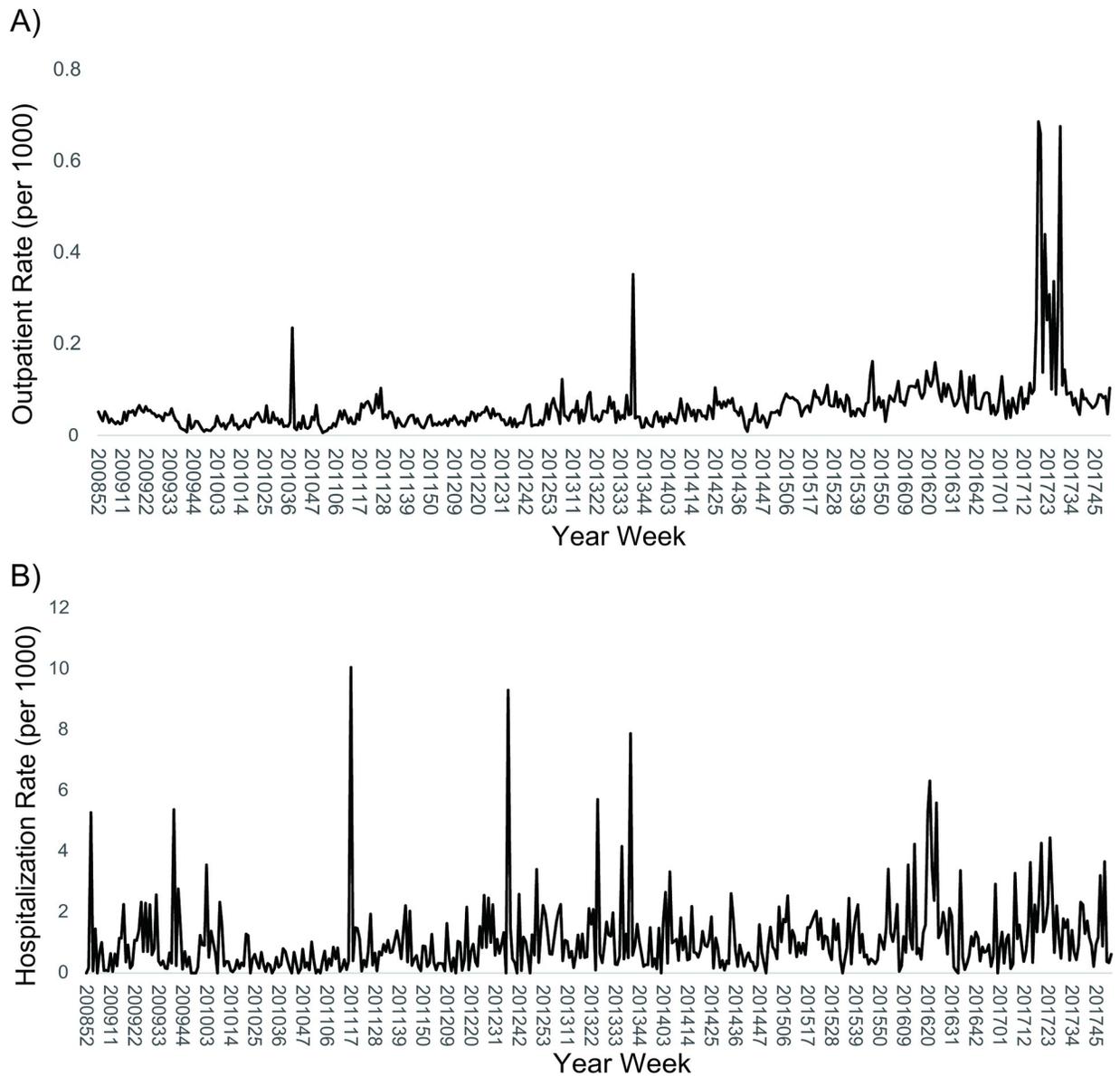
The temporal analysis showed that there were three peaks of scarlet fever outpatient cases during the study period. The first small peak occurred in October 2010, the second peak (2013) also began in October, and the highest level was from May to July 2017 (Fig 1A). The average hospitalization rate usually peaked in April, June and September (Fig 1B). The highest three hospitalization rates occurred in April 2011, September 2012 and September 2013.

**Table 2. Descriptive statistics on annual rate of scarlet fever in inpatient setting from 2009 to 2017 and annual percentage change (APC).**

Descriptive Statistics						
Hospitalization Age group	N1	Rate Range	Mean (per 1,000)	Std. Deviation	N2	APC% (95%CI)
0–2	198	(0, 3.24)	0.5097	0.66963	9	3 (-2.7, 9.1)
3–4	198	(0, 55.56)	1.4694	4.23467	9	12.8 (7.2, 18.6)
5–9	198	(0, 83.33)	2.5785	6.18071	9	16.0 (9.5, 22.9)
10–14	198	(0, 7.84)	0.6785	1.21436	9	16.3 (9.4, 23.4)
15+	198	(0, 0.05)	0.002	0.00695	9	8 (-7.5, 26.2)
Overall	198	(0, 4.02)	0.1265	0.32537	9	7.7 (4.5, 10.9)

Abbreviations: CI, confidence interval. Age is in years. N1 represents annual rate in 22 cities or counties for 9 years. N2 represents annual rate in the whole of Taiwan for 9 years.

<https://doi.org/10.1371/journal.pone.0215434.t002>



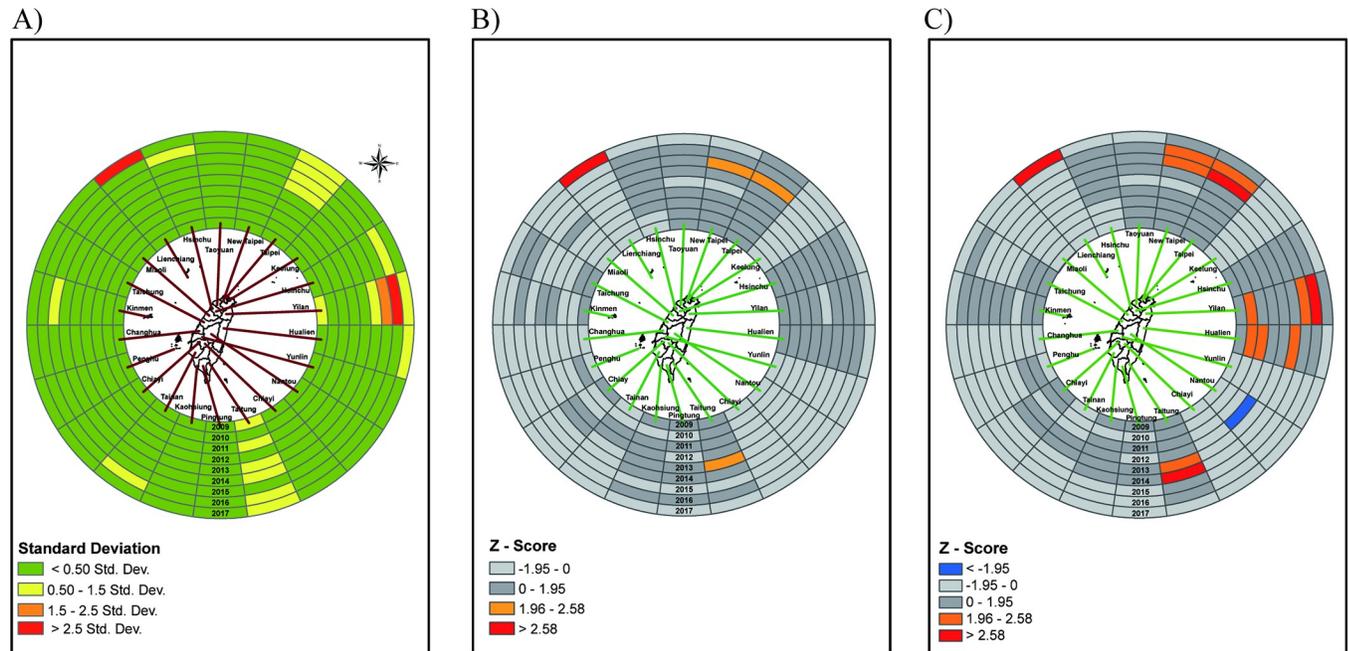
**Fig 1. Temporal trend of scarlet fever, 2009 to 2017.** (A) Outpatient visit rate. (B) Hospitalization rate.

<https://doi.org/10.1371/journal.pone.0215434.g001>

### Spatio-temporal analysis

The ring map of the outpatient visit rate is presented as a  $G_i^*$  result for patients 3 to 4 years of age. Fig 2A shows Z scores of standard deviation in outpatient visits by four age groups. We use Z scores to measure the distance from the mean in standard deviations; under the normal distribution, scores within one standard deviation from the average value account for 68.27% of the data, those within two standard deviations reach 95.45%, and those within three standard deviations reach 99.73%. The red color in the map means a Z score higher than 2.5 standard deviations away from the mean. Lienchiang County (upper left corner) is the primary hot spot, with a Z score greater than 2.5 standard deviations. Yilan County, Kinmen County and Taipei City are the secondary hotspots, with Z scores between 1.5 to 2.5 standard deviations.





**Fig 3. Ring map of outpatient visit rate of scarlet fever in age group 5–9 years.** (A) A Z score ring map from innermost in 2009 to outermost in 2017. (B) A ring map of Z-scores by applying spatio-temporal  $G_i^*$  without time lag. (C) A ring map of Z-scores by applying spatial-temporal  $G_i^*$  with 1 year lag.

<https://doi.org/10.1371/journal.pone.0215434.g003>

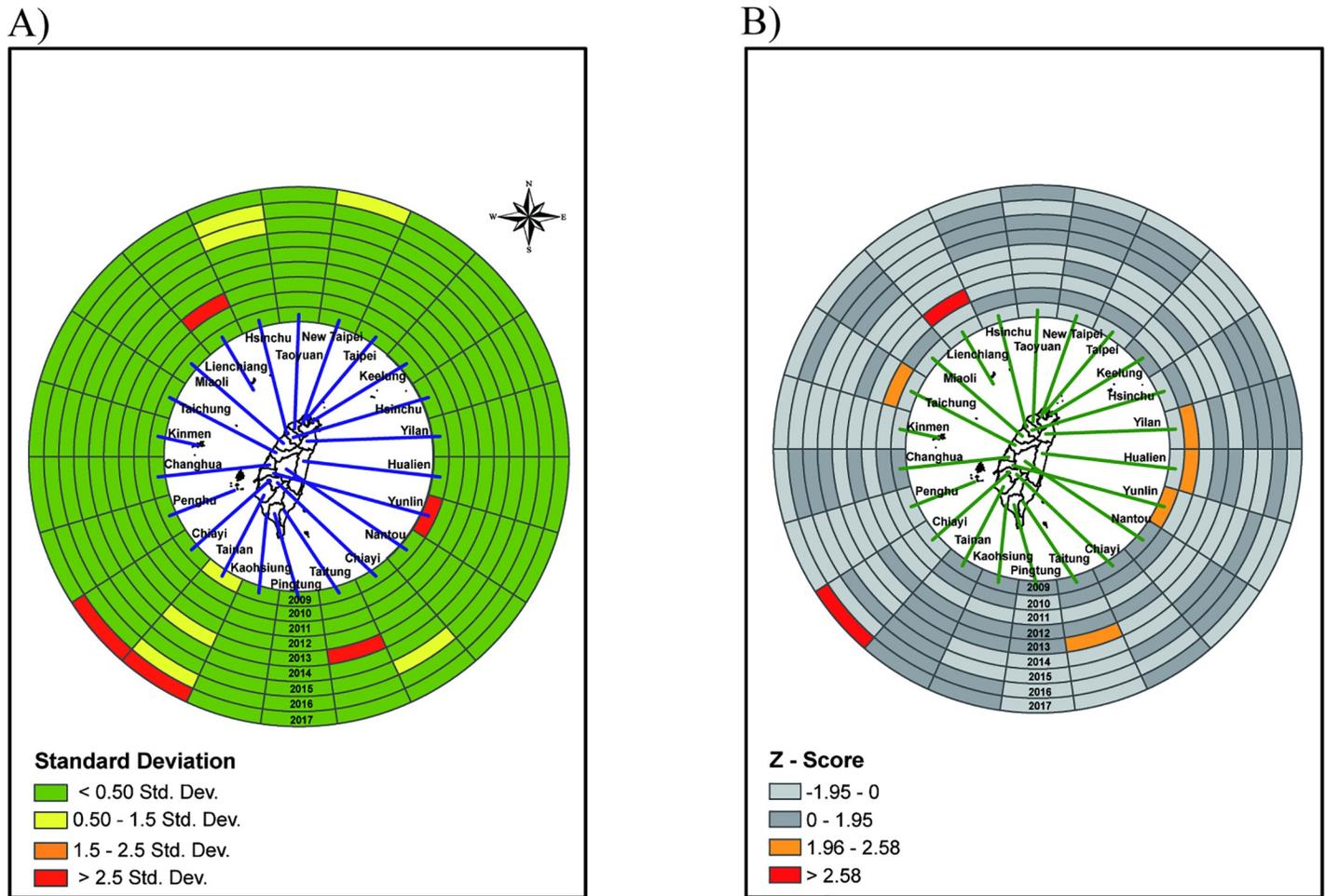
significant difference between the spatio-temporal  $G_i^*$  with or without time lag in the ring map (Fig 5B).

### Discussion

In this article, we have proposed a new spatio-temporal  $G_i^*$  statistic to cope with the question associated with the Gettis-Ord  $G_i^*$  statistic where the time-to-time autocorrelation of spatio-temporal data could not be taken into account in hotspot detection. Many hotspot detection approaches first perform a spatial characterization of the data then find the temporal pattern. The work presented in this paper sets itself apart from other studies by finding temporal intervals in the dataset. The temporal neighborhood was defined as a certain time window which is determined by the number of consecutive time-lagged significant correlation coefficients. Then a modified weighted function in the proposed spatio-temporal  $G_i^*$  statistic was used to precisely reflect the correlation between space and time in the data.

Mapping the distribution of diseases is a common exercise, and time and space are important factors which affect the spread of infectious diseases. Therefore, in addition to visualizing the outpatient visit rates and hospitalization rates for scarlet fever with a ring map GIS package, this study uses spatio-temporal  $G_i^*$  to present spatial and temporal scarlet fever hotspots. The information helps us understand where the hotspots are, information which needs to be taken into account by both scholars and policy makers.

We have provided a new approach to show scarlet fever hotspots in Taiwan by using ring maps and a novel spatio-temporal  $G_i^*$  statistic. The annual percentage changes in outpatient visit rate and hospitalization rate were 15.1% and 7.7% during 2009 to 2014. The highest two infected groups were children aged 3–5 and 5–9 years in our research. The main epidemic peaks were from March to May, and the high-risk population was 5–6 years old [32]. Likewise, a previous study by Mahara, et al. [5] reported that the average incidence (83.8%) was among



**Fig 4. Ring map of hospitalization rate of scarlet fever in age group 3–4 years.** (A) A Z score ring map from innermost in 2009 to outermost in 2017. (B) A ring map of Z-scores by applying spatio-temporal  $G_i^*$  without time lag.

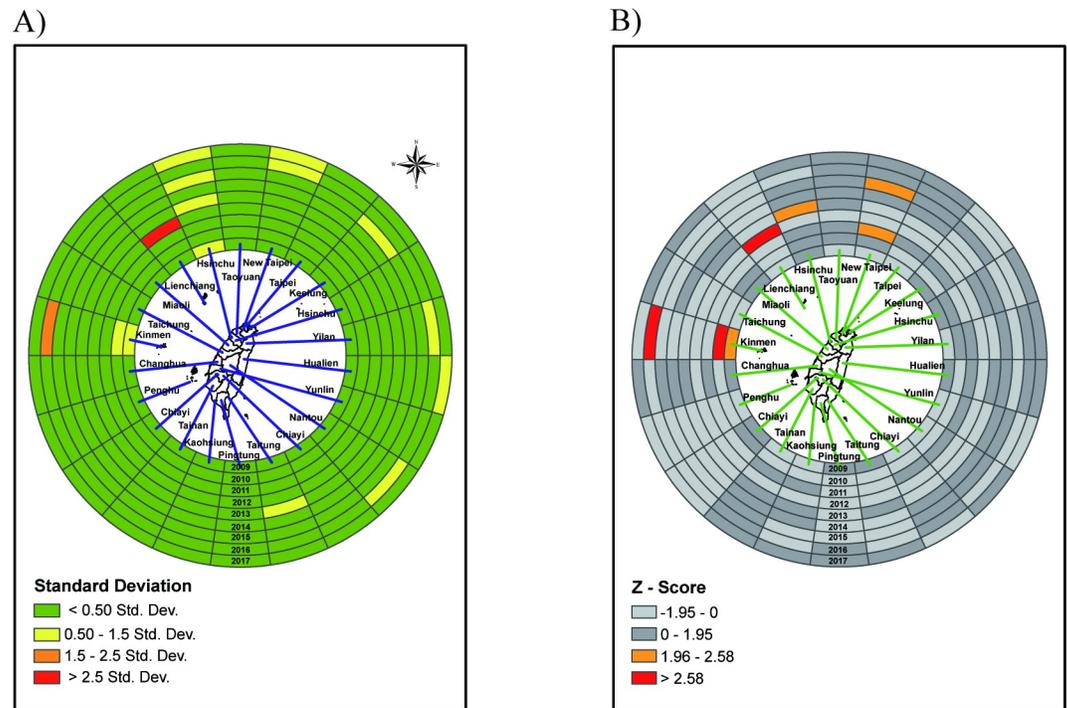
<https://doi.org/10.1371/journal.pone.0215434.g004>

children 3–8 years old. Additionally, several studies in China (including Hong Kong) and South Korea found a higher incidence among children 4–7 years of age. Another study in the United Kingdom reported that 87% of cases were in children under 10 years of age, while the Department of Health, UK found that scarlet fever was most common between the ages of 2 and 8 years [12, 33].

The temporal analysis in our study found that outpatient visits peaked in September and May, while the peak of hospitalization occurred in April and October. Similar findings were reported by Wu et al. and Liang et al. in Taiwan, who showed that the peaks occurred in March to July [14, 32]. Incidence has also been reported to peak from March to June with a small second peak from November to January in China [34], from March to May in the UK [35], and during winter months in Korea [36].

The seasonal peak has varied in different regions, which might be attributable to climate, population density, geographic distribution and other public health factors [37]. As shown above, transmission of scarlet fever tends to occur more in spring and winter compared to other seasons.

A previous study in Taiwan had similar results from 2000 to 2006, and also reported that the cases of scarlet fever in Hualien and Taitung Counties in eastern Taiwan increased



**Fig 5. Ring map of hospitalization rate of scarlet fever in age group 5–9 years.** (A) A Z score ring map from innermost in 2009 to outermost in 2017. (B) A ring map of Z-scores by applying spatio-temporal  $G_i^*$  without time lag.

<https://doi.org/10.1371/journal.pone.0215434.g005>

annually during that period [32]. In our study, the main epidemic peak was from March to May. Also, based on the percentage of outpatient visits in our study, cases in Yilan, Hualien and Taitung Counties are still increasing yearly from 2009 to 2017. Based on the original hospitalization rates, cases in Yilan and Hualien Counties (eastern Taiwan) were high, and the epidemic in Tainan City (western Taiwan) was even more pronounced. Scarlet fever hotspots occurred not only in eastern but also in western Taiwan and surrounding islands. Convenient transportation obviously lowers the geographical barriers to the spread of this infectious disease.

During our study period, scarlet fever, measured both in outpatient visits and the hospitalization rate, increased year by year. Although scarlet fever can be treated with drugs, serious complications such as toxic shock syndrome may occur if drugs are not used in time. In addition, sepsis and death are serious outcomes which bring great disease burden [38]. Nowadays, scarlet fever has been removed from Taiwan’s notifiable infectious diseases list. However, scarlet fever cases have resurged in the UK and China (including Hong Kong) in recent years, without any real cause being found [39]. By monitoring the outpatient and hospitalization rates of scarlet fever in Taiwan, we also found a significant increasing trend in recent years. It is recommended that the genetic evolution of group A *Streptococcus* (GAS) and multi-drug resistance be closely monitored [3].

The proposed spatio-temporal  $G_i^*$  statistic was illustrated with yearly county-level data from scarlet fever cases in Taiwan during 2009 to 2017. The retrospective space-time scan statistic by SaTScan (<https://www.satscan.org/>) was applied for age groups 3–4 and 5–9 years, and results are shown in the Supplement (S5–S8 Figs). Compared with the proposed spatio-temporal  $G_i^*$  statistic, the retrospective space-time scan statistic in the age group 3–4 and 5–9 years has fewer noticeable hotspots and clusters.

In addition, computing time depends highly on the size of the dataset. With our data, the calculation procedure takes 20 seconds to run on an Intel Core i5-8400 (2.8 GHz) computer. The data used in this study have been aggregated from the record level to county level, which ensures the identities of individuals cannot be determined by a reasonably foreseeable method. For record-level data or a real-time disease surveillance system, a space-time domain decomposition approach for parallel computation can reduce the effort required to identify hotspot patterns [40].

All spatial clustering approaches, regardless of their theoretical underpinning, statistical foundation, or mathematical specification, have limitations in accuracy, sensitivity, and the computational effort required for identifying clusters. As a result, a major challenge in practice is determining which technique will provide the most meaningful insights for a particular issue or dataset. The analytical results in this study have two limitations. First, our results are sensitive to spatio-temporal neighborhood size. As neighborhood size increases, hotspots will become larger and fewer; smaller neighborhood sizes capture more localized trends. There is inevitably an element of subjectivity in choosing an appropriate size for both spatial and temporal neighborhoods. To determine the appropriate size for both spatial and temporal neighborhoods, systematic data quality checks and analytic adjustments are needed. The choice of spatial neighborhood should reflect inherent relationships or characteristics of data. Once the spatial neighborhood is designated, the temporal neighborhood is constructed by discovering the similarity of autocorrelation. Second, scarlet fever has not been listed as a notifiable infectious disease in Taiwan since 2007, so some mild cases may have gone unreported.

## Conclusion

A spatial-temporal  $G_i^*$  statistic was proposed in this paper to detect hotspots in the space-time domain. A ring map was used to summarize the Z scores calculated by the spatial-temporal  $G_i^*$  statistic and to present an array of regional attributes in a single spatio-temporal reference graphic.

In conclusion, our study provides a new approach for better understanding of the spatio-temporal patterns of scarlet fever in Taiwan from 2009 to 2017. First, the increases in hotspots were mostly distributed in rural and offshore islands, especially in Hualien, Lienchiang and Jinmen Counties. For offshore islands, scarlet fever may be imported from China because of frequent contact and geographic proximity. Second, children under 9 years of age were more sensitive to scarlet fever infection, especially those aged 5–9. With the help of the spatio-temporal  $G_i^*$  statistic, public health officials can identify the high risk areas, which will help them monitor scarlet fever epidemics more effectively and reduce the disease burden among children.

## Supporting information

**S1 Fig. Annual outpatient visit rate from innermost ring in 2009 to outermost ring in 2017 for age 3–4.**

(DOCX)

**S2 Fig. Annual outpatient visit rate from innermost ring in 2009 to outermost ring in 2017 for age 5–9.**

(DOCX)

**S3 Fig. Annual hospitalization rate from innermost ring in 2009 to outermost ring in 2017 for age 3–4.**

(DOCX)

**S4 Fig. Annual hospitalization rate from innermost ring in 2009 to outermost ring in 2017 for age 5–9.**

(DOCX)

**S5 Fig. Annual outpatient visit rate from innermost ring in 2009 to outermost ring in 2017 by SaTScan for age 3–4.**

(DOCX)

**S6 Fig. Annual outpatient visit rate from innermost ring in 2009 to outermost ring in 2017 by SaTScan for age 5–9.**

(DOCX)

**S7 Fig. Annual hospitalization rate from innermost ring in 2009 to outermost ring in 2017 by SaTScan for age 3–4.**

(DOCX)

**S8 Fig. Annual hospitalization rate from innermost ring in 2009 to outermost ring in 2017 by SaTScan for age 5–9.**

(DOCX)

**S1 Table. The lag selection of outpatient visits.**

(DOCX)

**S2 Table. The lag selection of hospitalizations.**

(DOCX)

## Acknowledgments

This research was supported by a grant titled "Multidisciplinary Health Cloud Research Program: Technology Development and Application of Big Health Data" from Academia Sinica. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author Contributions

**Conceptualization:** Ta-Chien Chan.

**Formal analysis:** Jia-Hong Tang, Tzu-Jung Tseng.

**Methodology:** Jia-Hong Tang, Ta-Chien Chan.

**Resources:** Ta-Chien Chan.

**Supervision:** Ta-Chien Chan.

**Visualization:** Tzu-Jung Tseng, Ta-Chien Chan.

**Writing – original draft:** Jia-Hong Tang, Tzu-Jung Tseng.

**Writing – review & editing:** Ta-Chien Chan.

## References

1. Chiou CS, Liao TL, Wang TH, Chang HL, Liao JC, Li CC. Epidemiology and molecular characterization of *Streptococcus pyogenes* recovered from scarlet fever patients in central Taiwan from 1996 to 1999. *J Clin Microbiol.* 2004; 42(9):3998–4006. <https://doi.org/10.1128/JCM.42.9.3998-4006.2004> PMID: 15364982
2. Cunningham MW. Pathogenesis of group A streptococcal infections. *Clin Microbiol Rev.* 2000; 13(3):470–511. PMID: 10885988

3. Ben Zakour NL, Davies MR, You YH, Chen JHK, Forde BM, Stanton-Cook M, et al. Transfer of scarlet fever-associated elements into the group A *Streptococcus* M1T1 clone. *Scientific Reports*. 2015; 5.
4. Kim JH, Cheong HK. Increasing number of scarlet fever cases, South Korea, 2011–2016. *Emerging Infectious Diseases*. 2018; 24(1):172–3. <https://doi.org/10.3201/eid2401.171027> PMID: 29260684
5. Mahara G, Wang C, Huo D, Xu Q, Huang F, Tao L, et al. Spatiotemporal pattern analysis of scarlet fever incidence in Beijing, China, 2005–2014. *Int J Environ Res Public Health*. 2016; 13(1). Epub 2016/01/20. <https://doi.org/10.3390/ijerph13010131> PMID: 26784213
6. Zhang Q, Liu WD, Ma W, Shi YY, Wu Y, Li Y, et al. Spatiotemporal epidemiology of scarlet fever in Jiangsu Province, China, 2005–2015. *BMC Infectious Diseases*. 2017; 17.
7. Liu Y, Chan TC, Yap LW, Luo Y, Xu W, Qin S, et al. Resurgence of scarlet fever in China: a 13-year population-based surveillance study. *Lancet Infect Dis*. 2018. [https://doi.org/10.1016/S1473-3099\(18\)30231-7](https://doi.org/10.1016/S1473-3099(18)30231-7) PMID: 29858148
8. Hsieh YC, Huang YC. Scarlet fever outbreak in Hong Kong, 2011. *Journal of Microbiology Immunology and Infection*. 2011; 44(6):409–11.
9. Luk EYY, Lo JYC, Li AZL, Lau MCK, Cheung TKM, Wong AYM, et al. Scarlet fever epidemic, Hong Kong, 2011. *Emerging Infectious Diseases*. 2012; 18(10):1658–61. <https://doi.org/10.3201/eid1810.111900> PMID: 23018120
10. Feeney KT, Dowse GK, Keil AD, Mackaay C, McLellan D. Epidemiological features and control of an outbreak of scarlet fever in a Perth primary school. *Commun Dis Intell Q Rep*. 2005; 29(4):386–90. PMID: 16465930
11. Staszewska E, Kondej B, Czarkowski MP. Scarlet fever in Poland in 2012. *Przegl Epidemiol*. 2014; 68(2):209–12, 329–31. PMID: 25135501
12. Guy R, Williams C, Irvine N, Reynolds A, Coelho J, Saliba V, et al. Increase in scarlet fever notifications in the United Kingdom, 2013/2014. *Euro Surveill*. 2014; 19(12):20749. Epub 2014/04/05. PMID: 24698137
13. Basetti S, Hodgson J, Rawson TM, Majeed A. Scarlet fever: a guide for general practitioners. *London J Prim Care (Abingdon)*. 2017; 9(5):77–9. <https://doi.org/10.1080/17571472.2017.1365677> PMID: 29081840
14. Wu W-J, Liu Y-L, Kuo H-W, Huang W-T, Yang S-L, Chuang J-H. An integrated syndromic surveillance system for monitoring scarlet fever in Taiwan. *Online Journal of Public Health Informatics*. 2013; 5(1): e160.
15. Goovaerts P, Jacquez GM. Detection of temporal changes in the spatial distribution of cancer rates using local Moran's I and geostatistically simulated spatial neutral models. *J Geogr Syst*. 2005; 7(1):137–59. <https://doi.org/10.1007/s10109-005-0154-7> PMID: 16710441
16. Tango T. *Statistical methods for disease clustering*. New York: Springer; 2010.
17. Tango T, Takahashi K, Kohriyama K. A space-time scan statistic for detecting emerging outbreaks. *Biometrics*. 2011; 67(1):106–15. <https://doi.org/10.1111/j.1541-0420.2010.01412.x> PMID: 20374242
18. Pickle LW, Mungiole M, Jones GK, White AA. Exploring spatial patterns of mortality: the new atlas of United States mortality. *Stat Med*. 1999; 18(23):3211–20. PMID: 10602146
19. Doi Y, Yokoyama T, Tango T, Takahashi K, Fujimoto K, Nakano I. Temporal trends and geographic clusters of mortality from amyotrophic lateral sclerosis in Japan, 1995–2004. *J Neurol Sci*. 2010; 298(1–2):78–84. <https://doi.org/10.1016/j.jns.2010.08.004> PMID: 20804988
20. Koch T. *Cartographies of disease: maps, mapping, and medicine*. 1st ed. Redlands, Calif.: ESRI Press; 2005.
21. Battersby SE, Stewart JE, Lopez-De Fede A, Remington KC, Mayfield-Smith K. Ring maps for spatial visualization of multivariate epidemiological data. *Journal of Maps*. 2011; 7(1):564–72.
22. Speckmann B, Verbeek K. Necklace maps. *IEEE Trans Vis Comput Graph*. 2010; 16(6):881–9. <https://doi.org/10.1109/TVCG.2010.180> PMID: 20975124
23. Anselin L. Local indicators of spatial association—Lisa. *Geographical Analysis*. 1995; 27(2):93–115.
24. Getis A, Ord JK. The analysis of spatial association by use of distance statistics. *Geographical Analysis*. 1992; 24(3):189–206.
25. Ord JK, Getis A. Local spatial autocorrelation statistics—distributional issues and an application. *Geographical Analysis*. 1995; 27(4):286–306.
26. Waller LA, Gotway CA. *Applied spatial statistics for public health data*. Hoboken, N.J.: John Wiley & Sons; 2004.
27. Chan Ta-Chien, Wang Chien-Min, Lee Yung-Mei. Looking at temporal changes—use this Python tool for creating ring maps. *ArcUser*. 2013:1–3.

28. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for Joinpoint regression with applications to cancer rates. *Stat Med*. 2000; 19(3):335–51. Epub 2000/01/29. PMID: [10649300](#)
29. Bhunia GS, Kesari S, Chatterjee N, Kumar V, Das P. Spatial and temporal variation and hotspot detection of kala-azar disease in Vaishali district (Bihar), India. *BMC Infect Dis*. 2013; 13:64. <https://doi.org/10.1186/1471-2334-13-64> PMID: [23375077](#)
30. Stopka TJ, Goulart MA, Meyers DJ, Hutcheson M, Barton K, Onofrey S, et al. Identifying and characterizing hepatitis C virus hotspots in Massachusetts: a spatial epidemiological approach. *BMC Infect Dis*. 2017; 17(1):294. <https://doi.org/10.1186/s12879-017-2400-2> PMID: [28427355](#)
31. Plant RE. *Spatial data analysis in ecology and agriculture using R*. Boca Raton: CRC Press; 2012.
32. Liang J-H, Lin Y-L, Ke J-F, Li C-F, Li Y-C, Shy W-Y. Epidemiologic analysis of scarlet fever reported from eastern Taiwan during 2000–2006. *Epidemiol Bull (Taipei Taiwan)*. 2010; 26 (20):354–63.
33. Department of Health U. Scarlet fever: second year of high activity 2015 [cited 2018 5/26]. Available from: <https://www.gov.uk/government/news/scarlet-fever-second-year-of-high-activity>.
34. Duan Y, Huang XL, Wang YJ, Zhang JQ, Zhang Q, Dang YW, et al. Impact of meteorological changes on the incidence of scarlet fever in Hefei City, China. *Int J Biometeorol*. 2016; 60(10):1543–50. Epub 2016/03/05. <https://doi.org/10.1007/s00484-016-1145-8> PMID: [26932715](#)
35. Turner CE, Pyzio M, Song B, Lamagni T, Meltzer M, Chow JY, et al. Scarlet fever upsurge in England and molecular-genetic analysis in north-west London, 2014. *Emerg Infect Dis*. 2016; 22(6):1075–8. Epub 2016/05/19. <https://doi.org/10.3201/eid2206.151726> PMID: [27192393](#)
36. Park DW, Kim SH, Park JW, Kim MJ, Cho SJ, Park HJ, et al. Incidence and characteristics of scarlet fever, south Korea, 2008–2015. *Emerg Infect Dis*. 2017; 23(4):658–61. Epub 2017/03/23. <https://doi.org/10.3201/eid2304.160773> PMID: [28322696](#)
37. Mahara G, Wang C, Yang K, Chen S, Guo J, Gao Q, et al. The association between environmental factors and scarlet fever incidence in Beijing region: using GIS and spatial regression models. *Int J Environ Res Public Health*. 2016; 13(11). Epub 2016/11/10. <https://doi.org/10.3390/ijerph13111083> PMID: [27827946](#)
38. Kissoon N, Uyeki TM. Sepsis and the global burden of disease in children. *JAMA Pediatr*. 2016; 170(2):107–8. <https://doi.org/10.1001/jamapediatrics.2015.3241> PMID: [26661465](#)
39. Lamagni T, Guy R, Chand M, Henderson KL, Chalker V, Lewis J, et al. Resurgence of scarlet fever in England, 2014–16: a population-based surveillance study. *Lancet Infect Dis*. 2018; 18(2):180–7. Epub 2017/12/02. [https://doi.org/10.1016/S1473-3099\(17\)30693-X](https://doi.org/10.1016/S1473-3099(17)30693-X) PMID: [29191628](#)
40. Hohl A, Delmelle E, Tang W, Casas I. Accelerating the discovery of space-time patterns of infectious diseases using parallel computing. *Spat Spatiotemporal Epidemiol*. 2016; 19:10–20. <https://doi.org/10.1016/j.sste.2016.05.002> PMID: [27839573](#)