Brazilian Journal of Biometrics, **42**, 329–338 (2024)
doi:10.28951/bjb.v42i4.707

BRAZILIAN JOURNAL OF BIOMETRICS
ISSN:2764-5290

ARTICLE

Application of spatio-temporal scan statistics in cases of intentional homicides in northern Brazil

Henrique José de Paula Alves*[,1] and Ben Dêivide de Oliveira Batista[2]

[1] Institute of Applied Economic Research, Rio de Janeiro, RJ, Brazil
[2]Federal University of Sao João del-Rei, Sao João del-Rei, MG, Brazil
*Corresponding author. Email: jpahenrique@gmail.com

(Received: December 13, 2023; Revised: May 10, 2024; Accepted: May 17, 2024; Published: October 15, 2024)

Abstract

The crime of intentional homicide in Brazil is worrying. In the northern region, this type of crime has been growing since 2020. In this sense, we decided to apply Kuldorff's prospective space-time scan statistics in order to identify emerging clusters. We conclude that there are two emerging clusters of high relative risk (Amazonas and Pará) that require rapid intervention and two clusters of low relative risk (Acre, Roraima, Amazonas, Amapá, and Pará) that do not require urgent intervention. Some characteristics of these two clusters are presented: radius, population, relative risk, likelihood ratio.

Keywords: Likelihood ratio test, scan statistics, spatio-temporal clusters.

1. Introduction

Violence is a serious problem and constitutes a challenge for the health sector due to its impact and negative repercussions in all areas of life, both individual and collective. It is a multifactorial phenomenon related to historical, cultural, contextual, structural and interpersonal factors (Leite *et al.,* 2017). Violent actions are not only related to physical force, but also to attitudes that compromise the emotional and psychological aspects of the victim or even any other way that can cause suffering to the individual and their family (Dim, 2021).

Homicide is characterized as the the most extreme manifestation of interpersonal violence, being the main cause of death (Tavares *et al.,* 2016). This phenomenon generates potential losses within the scope of public policies and highlights structural deficiencies in economic and social development, imposing the need for effective strategies to prevent and reduce mortality rates. Considering that homicide is a consequence of interpersonal violence, it is observed that more people die as a result of this problem worldwide than in all other criminal types since the year 2000.

According to the Atlas of Violence 2023, an important document that reports violence in Brazil, published annually by the Institute of Applied Economic Research (Ipea), in Brazil there were 616, 095 victims of intentional homicides between the years 2011 and 2021 and, in that same period, there was a drop in the rate per 100, 000 inhabitants of 18.3%. Of these, 77% are afro-Brazilian people victims and, among indigenous people, the rate per 100, 000 inhabitants increased by 29% in the same period. The risk of a afro-Brazilian people person being murdered is 2.9 times greater when compared to a non-afro-Brazilian people individual. If there had not been an increase in firearms in circulation between 2019 and 2021, Brazil would have recorded 6, 379 fewer homicides. These facts can be consulted at https://www.ipea.gov.br/atlasviolencia/arquivos/artigos/7868-atlasviolencia-2024-v11.pdf.

These high numbers occur due to existing problems in society, such as inequality, organized crime, impunity, corruption and private living spaces that favor domestic violence. Among the regions of Brazil, according to the Atlas of Violence 2023, the northeast and northern regions of Brazil have the highest rates of intentional homicides per 100, 000 inhabitants. The northern region, from 2020 onwards, is the most worrying, with rates even higher than Brazil itself. Several factors are associated with homicides, such as psychological or biological dysfunctions at the individual level, or factors criminogens such as weapons and psychotropic drugs, which can influence interpersonal conflicts and the use of lethal violence.

Social, demographic and economic elements, for example, income, socioeconomic inequality, population concentration and age structure, are conditions for the actions of individuals (Tavares *et al.,* 2016). In addition to the factors already mentioned, it is possible to say that the occurrence of the problem may be linked to the exploitation of the riches of land located in the northern region of Brazil, since this almost entirely covers the Brazilian Amazon forest. Uncontrolled logging, fishing and mining and other factors contribute to the increase in violent deaths in this region. Detecting the counties in the northern region that have a higher relative risk of victims of intentional homicides is essential.

These reasons are sufficient for a study aimed at the northern region of Brazil to combat this type of violent crime to be carried out. The analysis and evaluation of spatial and space-time patterns and trends in intentional homicides are necessary to control and eliminate these homicides.

Kulldorff's spatial-temporal scan statistical method can detect the distribution characteristics in the temporal and spatial axes, which is consistent with real-world conditions (Kulldorff & Nagarwalla, 1995; Qian *et al.,* 2016; Vieira *et al.,* 2014; Xia *et al.,* 2015). Furthermore, the relative risk of any phenomenon in a cluster area can be estimated by comparing it with the area outside the cluster area. For this reason, this method has been widely used in homicide studies (Mohammadi, 2023; Silva *et al.,* 2023; Steelesmith & Lindstrom, 2023; Uittenbogaard & Ceccato, 2012).

In this study, our objective was to use the Kulldorff scan statistical analysis to explore the prospective spatio-temporal dynamics of intentional homicides at the level of counties in the northern Brazilian region, since this region presents itself as the most violent in Brazil in the last three years.

2. Material

The northern Brazilian region is known worldwide for having the largest tropical forest in the world: the Amazon forest. Extensive varieties of fauna and flora are found. Mining, deforestation and illegal fishing are recurrent in this region and, in this sense, territorial disputes with indigenous civilizations are a strong justification for the considerable increase in violent deaths in this region in the last three years. Figure 1 highlights the northern region on the map of Brazil and presents the territorial distribution of indigenous conservation areas, locations with easy access to healthcare and also located schools.

These characteristics form an important contributing factor to the increase in violent deaths in the northern region of Brazil. Table 1 describes the federation units belonging to the northern

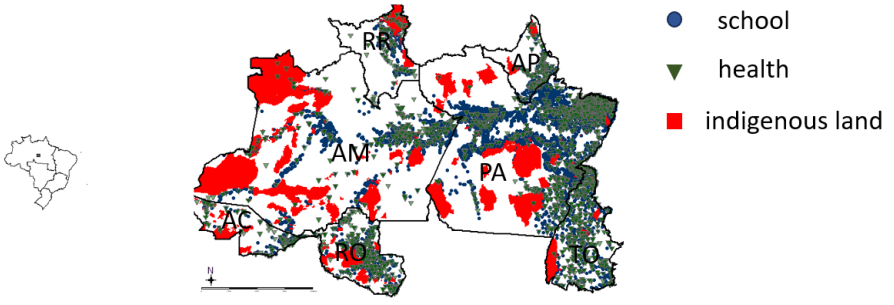


Figure 1. Characterization of the northern region with the demarcation of indigenous land, school areas and areas with easy access to health care.

Brazilian region. The state of Pará (PA) has the largest number of counties (144) and is also the most populous, also presenting the largest number of victims of intentional homicide over the period between 2018 and 2022. Roraima (RR) is the less populous state and with fewer victims of intentional homicide over the period between 2018 and 2022.

Table 1. Description of federative unit in the northern Brazilian region - 2018 to 2022. A total of 450 counties were considered

FU	Counties	Population (victims)				
		2018	2019	2020	2021	2022
RO	52	1,757,589 (395)	1,777,225 (371)	1,796,460 (403)	1,815,278 (416)	1,815,278 (479)
AC	22	869,265 (396)	881,935 (270)	894,470 (270)	906,876 (171)	906,876 (207)
AM	62	4,080,611 (1,101)	4,144,597 (996)	4,207,714 (923)	4,269,995 (1,464)	4,269,995 (1,319)
RR	15	576,568 (159)	605,761 (176)	631,181 (174)	652,713 (197)	652,713 (171)
PA	144	8,513,497 (3,827)	8,602,865 (2,764)	8,690,745 (2,198)	8,777,124 (2,193)	8,777,124 (2,185)
AP	16	829,494 (301)	845,731 (232)	861,773 (239)	877,613 (281)	877,613 (210)
TO	139	1,555,229 (368)	1,572,866 (336)	1,590,248 (404)	1,607,363 (338)	1,607,363 (396)

FU: represents the acronym of the federative unit. Counties: is the number of counties in each FU. Here: RO - Rondônia, AC - Acre, AM - Amazonas, RR - Roraima, PA - Pará, AP - Amapá, TO - Tocantins.

Homicide rates are important indicators for measuring and comparing the situation of this important crime in the FU's belonging to the northern region. Because we believe that the calculation of rates per 100, 000 inhabitants is quite widespread in the literature, we chose not to highlight how to obtain it here. Table 2 presents the rates per 100, 000 inhabitants for these municipalities. Rates are an important indicator of the presence of population heterogeneity in FU's (see Tabela 1).

Here, we highlight Acre (AC) which had the highest rate of intentional homicide in 2019 (60.61) and Roraima (RO) which, in the same year, had the lowest rate (20.87). We also note that Acre (AC)

Table 2. Description of rates of federative unit in the northern Brazilian region - 2018 to 2022. A total of 450 counties were considered

FU	Counties	Rate (100.000)				
		2018	2019	2020	2021	2022
RO	52	22.47	20.87	22.43	22.91	26.39
AC	22	45.55	60.61	30.18	18.85	22.82
AM	62	26.98	24.03	21.93	34.28	30.89
RR	15	27.57	29.05	27.57	30.18	26.20
PA	144	44.95	32.13	25.29	24.98	24.89
AP	16	36.28	27.43	27.73	32.02	23.93
TO	139	23.66	21.36	25.40	21.03	24.64

FU: represents the acronym of the federative unit. Counties: is the number of counties in each FU. Here: RO - Rondônia, AC - Acre, AM - Amazonas, RR - Roraima, PA - Pará, AP - Amapá, TO - Tocantins.

is the FU with the highest rates in almost all years except 2021 in which it presented the lowest rate among all FU's. Pará (PA) presents high rates in 2018 and 2019, showing a drop between 2018 and 2021, stabilizing in 2022.

2.1 The dataset

The data considered here are made available by the Brazilian National Public Security Secretariat (Senasp), and can be obtained at https://dados.mj.gov.br/dataset/sistema-nacional-de-estatisticas-de-seguranca-public. We prepare this data in such a way that information on the total number of victims in each counties in the northern region of Brazil is consolidated annually. The original database available has the number of intentional homicide victims consolidated monthly.

In a previous analysis and with unpublished results, we calculated the Local Moran Index in each year and, in the northern region, the clusters found present a strong correlation and are of the "high-high" type, suggesting a high correlation of intentional homicides in these areas (Coelho *et al.,* 2021). A limitation of the Local Moran Index is that it does not provide a measure of the risk of victims of this type of crime. We then decided to use the prospective spatio-temporal scan method, proposed by Martin Kulldorff (Kulldorff *et al.,* 1998), to obtain the relative risks of each of the clusters found.

3. Prospective spatio-temporal scan statistic

We used prospective spatio-temporal scan statistics according to a Poisson probability distribution model, which is implemented in the SaTScan TM software (Kulldorff, 2018; Kulldorff, 2001). This method finds the most likely clusters in several cylindrical candidates. The cylinders are determined by a circular basis (the spatial scan window) that is centered on a candidate location (centroid) and that has a radius r. The height of the cylinder is determined by the time scan window t. Therefore, each counties centroid is the center of a series of candidate groups for clusters of different radius r and heights t.

Consider the null hypothesis H_0: "There is no difference in the risk of intentional homicide between the inside and outside of the cylinder", and alternative hypothesis H_A: "There is a greater or lesser risk of intentional homicide inside the cylinder" . We calculate the number of expected cases μ using

$$\mu = p\frac{C}{P}, \tag{1}$$

where p is the population within the cylinder, C is the total number of intentional homicide victims and P is the total population. The most likely cluster is the one that maximizes the likelihood ratio given by the expression

$$LR = \frac{L(Z)}{L_0} = \frac{\left(\frac{n_z}{\mu(Z)}\right)^{n_z}\left(\frac{N-n_z}{N-\mu(Z)}\right)^{N-n_z}}{\left(\frac{N}{\mu(T)}\right)^{N}}, \tag{2}$$

where $L(Z)$ is the likelihood function for the cylindrical cluster candidate Z, L_0 is the likelihood function under H_0, n_z is the number of cases within the cylinder, $\mu(Z)$ is the number of cases expected in the cylinder Z, N is the number of cases observed for the entire study area during the entire study period T and $\mu(T)$ is the total number of cases expected in the study area over all time periods T.

High relative risk clusters are those in which the ratio LR is greater than 1, that is, $\frac{n_z}{\mu(Z)} > \frac{N-n_z}{N-\mu(z)}$. If the same ratio is less than 1, these clusters are of low relative risk. The LR ratio is a measure of how the relative risk inside a cylinder differs from the outside risk, and typically its logarithmic transformation is reported as the logarithmic likelihood ratio LLR.

We ran 999 Monte Carlo simulations for significance testing, where we randomized the locations and times of intentional homicide cases to obtain a likelihood ratio for each run and candidate cluster that form a distribution under H_0. Furthermore, we included statistically significant secondary clusters ($p < 0.05$) in our results. Finally, to illustrate the distribution of relative risk within the clusters, we report choropleth maps that contain relative risk $RR_{counties}$, which is the relative risk within a county divided by the relative risk outside that same county. The relative risk of each counties is given by the expression

$$RR_{countiess} = \frac{c_{counties}/e_{counties}}{(C - c_{counties})/(C - e_{counties})}, \tag{3}$$

where $c_{counties}$ is the total number of intentional homicides in a given counties, $e_{counties}$ is the number of intentional homicides expected in that counties, being obtained by the expression (1) using the population ($p_{counties}$) of the counties and C is the total number of intentional homicides observed.

In a similar way, we calculated the relative risk within each cluster according to the expression

$$RR_{cluster} = \frac{c_{cluster}/e_{cluster}}{(C - c_{cluster})/(C - e_{cluster})}, \tag{4}$$

where $c_{cluster}$, $e_{cluster}$ and C are, respectively, the total number of intentional homicides, the number of expected intentional homicides and the total number of intentional homicides observed within each cluster. To obtain e, we use the expression (1), with p being the cluster population.

Unlike its retrospective variant, prospective spatio-temporal scan statistics find clusters that are "active" or "emerging" at the end of the study period. Prospective analysis only evaluates cylinders that have a "current" end time. Therefore, the set of candidate cylinders is reduced to include only "active" cylinders at the end of the study period (Kulldorff, 2001), i.e., only those cylinders with an end date equal to the end of the study period.

4. Main results

The prospective spatio-temporal scan statistic did not detect significant clusters (5%) in the years 2018, 2019 and 2020. However, significant prospective spatio-temporal clusters of intentional homicides were detected in the period 2021-2022 (see Figure 2). The two low-risk clusters detected include, in cluster 1, the state of Acre (AC), Roraima (RO) and a large part of Amazonas (AM) and,

for Cluster 2, a large part of the state of Pará (PA) and some counties in Amapá (AP) are part of this cluster. For the two high-risk clusters, some counties in Amazonas and a large part of Pará form these clusters.

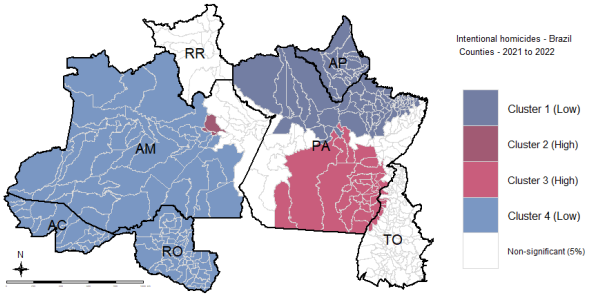


Figure 2. Clusters detected by prospective space-time scan statistics. Here, the term "High" means the cluster is high risk and the term "Low" means the cluster is low risk.

Among the clusters with the highest relative risk, although it has the smallest number of counties (2), cluster 2 is the most populous, has the highest relative risk and smallest coverage (57.59 km), with only 2 forming counties. Cluster 3 has 35 counties, has the smallest population, number of victims, relative risk, and has the largest coverage (381.43 km) (see Table 3).

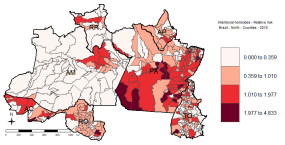
Table 3. Descriptive measures of significant high and low relative risk clusters detected

Level	Cluster	Measurements						
		Counties	Population	c (e)	RR	LLR	Radius (km)	p-value
Low	1	92	6,535,235	2,457 (3,631.9)	0.64	244.37	744.48	0.0234
	4	118	4,010,444	1,706 (2,297.87)	0.72	90.99	1,238.9	0.0421
High	2	2	2,260,868	2,144 (1,302.09)	1.70	241.63	57.59	0.0378
	3	35	1,526,900	1,338 (877,69)	1.55	108.03	381.43	0.0138

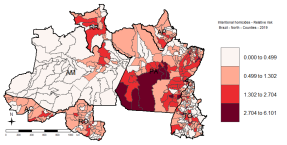
Here: c – total victims in the cluster, e – victims expected in the cluster, RR – relative risk, Radius – the radius of the circular base of the cluster, in kilometers, LLR – logarithmic likelihood ratio, p-value – indicates the significance of the clusters.

Table 3 details the characteristics of these high and low relative risk clusters, providing information such as number of counties, total number of intentional homicide victims, total cluster population, relative risk, logarithm of the likelihood ratio *LLR*. We note that, among the low relative risk clusters, cluster 1 is the most populous and has the highest number of victims, even though it contains the smallest number of counties (92) and the lowest relative risk (0.64), and is also the cluster of smaller scope (744.48 km). Cluster 4 is the least populous and contains the largest number of counties (118). This has the lowest number of victims, but with greater relative risk and greater coverage (1,238.9 km) (see Table 3).

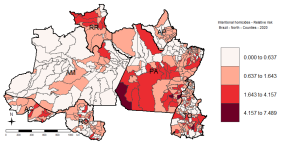
Figure 3 presents the relative risks of each counties in each of the years considered: 2018 to 2022.



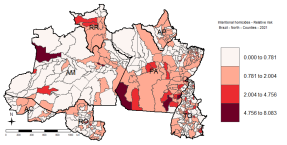
(a) year: 2018



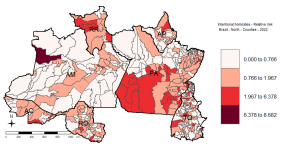
(b) year: 2019



(c) year: 2020



(d) year: 2021



(e) year: 2022

Figure 3. Relative risk (see expression (3)) in each of the counties forming the federative units in the northern region of Brazil.

Over the years, we noticed that the location of high relative risks is mainly concentrated in the south of the state of Pará (PA). In general, the FU of Pará and Amapá (AP) have the counties with the highest relative risks in all years (3).

5. Discussion

In this study, we carried out periodic surveillance of victims of intentional homicide using prospective spatio-temporal scan statistics, in counties in the northern region of Brazil. We detected emerging groups of victims at the municipal level and provided annual results during the study period from 2018 to 2022. Our objective is to show that this statistical tool is useful in guiding competent Brazilian authorities in directing resources and actions to combat this type of violence in Brazil, which is emerging and worrying mainly in the northern region. Additionally, we include informative numbers that illustrate changes in cluster characteristics as 2021 and 2022 unfold.

The periodic application of prospective spatio-temporal scan statistics is attractive because of its ability to consider updated case counts of intentional homicide victims on a daily, monthly, and even annual basis, as in the present study, while tracking previous existing clusters. Although it identifies emerging areas of concern (high relative risk clusters), this statistical method also identifies those areas that do not require immediate intervention (low relative risk clusters). However, the latter are important indicators of possible underreporting of cases. Furthermore, this method also makes it possible to track the characteristics of previously detected clusters, for example, to determine whether they are increasing or decreasing in magnitude, or whether the relative risk is increasing or not. These capabilities allow us to evaluate current strategies to control the spread of this type of crime and provide a basis for anticipating the future development of hot spots.

Despite the contributions of our study, there are a number of limitations and assumptions. First, we apply prospective spatio-temporal scan statistics in its basic form, which generates clusters with a circular shape base. Circles may be a poor choice in a study area that presents substantial spatial heterogeneity, causing overestimation or underestimation of significant clusters. Second, the number of confirmed intentional homicide victims largely depends on police efforts and records and, therefore, may not represent the true magnitude and spatial distribution of this type of crime. These are the well-known problems of under-registration. Third, clusters with irregular shapes may be undetected using circles as a basis (Costa *et al.,* 2012; Duczmal *et al.,* 2006; Liang *et al.,* 2020; Yiannakoulias *et al.,* 2007).

Even with all these limitations and assumptions presented, there are numerous studies with the aim of improving the performance of this method. What we did was show that the application of this, even in a simpler way, is valid in this type of study. Possible progress in this study is the possibility of using other relevant distributions in counting the analyzed cases, such as the negative binomial, binomial and normal distribution.

Acknowledgments

We would like to thank the two anonymous reviewers and the editor for their comments and suggestions which substantially improved the clarity of this article.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Coelho, D. S. C., Cunha, A. d. S., Alves, H. & Guedes, E. P. Metodologia para a pré-seleção dos municípios participantes do Programa Nacional de Enfrentamento de Homicídios e Roubos (2021).
2. Costa, M. A., Assunçao, R. M. & Kulldorff, M. Constrained spanning tree algorithms for irregularly-shaped spatial clustering. *Computational Statistics & Data Analysis* **56,** 1771–1783 (2012).
3. Dim, E. E. Experiences of physical and psychological violence against male victims in Canada: A qualitative study. *International journal of offender therapy and comparative criminology* **65,** 1029–1054 (2021).
4. Duczmal, L., Kulldorff, M. & Huang, L. Evaluation of spatial scan statistics for irregularly shaped clusters. *Journal of Computational and Graphical Statistics* **15,** 428–442 (2006).
5. Kulldorff, M. SaTScanTM user guide for version 9.6. 2018. *Department of Medicine, Harvard Medical School: Boston, MA* **2120** (2018).
6. Kulldorff, M. Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **164,** 61–72 (2001).
7. Kulldorff, M., Athas, W. F., Feurer, E. J., Miller, B. A. & Key, C. R. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *American journal of public health* **88,** 1377–1380 (1998).
8. Kulldorff, M. & Nagarwalla, N. Spatial disease clusters: detection and inference. *Statistics in medicine* **14,** 799–810 (1995).
9. Leite, F. M. C., Mascarello, K. C., Almeida, A. P. S. C., Fávero, J. L., Santos, A. S. d., Silva, I. C. M. d. & Wehrmeister, F. C. Analysis of the female mortality trend due to assault in Brazil, States and Regions. *Ciência & Saúde Coletiva* **22,** 2971–2978 (2017).
10. Liang, S., Han, D. & Yang, Y. Cluster validity index for irregular clustering results. *Applied Soft Computing* **95,** 106–583 (2020).
11. Mohammadi, A. Spatiotemporal patterns of homicide rates in Tehran metropolitan area, Iran. *Spatial Information Research,* 1–9 (2023).
12. Qian, H. *et al.* Detecting spatial-temporal cluster of hand foot and mouth disease in Beijing, China, 2009-2014. *BMC infectious diseases* **16,** 1–13 (2016).
13. Silva, C. C. d., Souza, K. O. C. d., Paz, W. S. d., Santos, A. P. S., Melo, L. R. S. d., Sousa, Á. F. L. d., Araújo, D. d. C. & Santos, A. D. d. Modelagem espacial da mortalidade por homicídios na Região Nordeste do Brasil. *Revista Brasileira de Enfermagem* **76,** e20220182 (2023).
14. Steelesmith, D. L. & Lindstrom, M. R. e. a. Spatiotemporal Patterns of Deaths of Despair Across the US, 2000-2019. *American journal of preventive medicine* (2023).
15. Tavares, R., Catalan, V. D. B., Romano, P. M. d. M. & Melo, E. M. Homicídios e vulnerabilidade social. *Ciência & Saúde Coletiva* **21,** 923–934 (2016).
16. Uittenbogaard, A. & Ceccato, V. Space-time clusters of crime in Stockholm, Sweden. *Rev. Eur. Stud.* (2012).
17. Vieira, C. P., Oliveira, A. M., Rodas, L. A. C., Dibo, M. R., Guirado, M. M. & Chiaravalloti Neto, F. Temporal, spatial and spatiotemporal analysis of the occurrence of visceral leishmaniasis in humans in the City of Birigui, State of São Paulo, from 1999 to 2012. *Revista da Sociedade Brasileira de Medicina Tropical* **47,** 350–358 (2014).

18. Xia, J. *et al.* Spatial, temporal, and spatiotemporal analysis of malaria in Hubei Province, China from 2004–2011. *Malaria journal* **14,** 1–10 (2015).

19. Yiannakoulias, N., Rosychuk, R. J. & Hodgson, J. Adaptations for finding irregularly shaped disease clusters. *International Journal of Health Geographics* **6,** 1–16 (2007).