

# Epidemiologia espacial e a detecção de aglomerados espaciais do dengue na Paraíba: uma comparação entre os métodos *Scan* flexível e *Scan* circular

## *Spatial epidemiology and the detection of spatial clusters of dengue in Paraíba: a comparison between the flexible Scan and circular Scan methods*

Ana Cláudia Oliveira de Melo<sup>1</sup> , José Carlos da Silva Melo<sup>1</sup> , Ronei Moraes<sup>1</sup> 

<sup>1</sup> Departamento de Estatística, Centro de Ciências Exatas e da Natureza (CCEN), Universidade Federal da Paraíba (UFPB) - João Pessoa (PB), Brasil.

**Como citar:** Melo ACO, Melo JCS, Moraes R. Epidemiologia espacial e a detecção de aglomerados espaciais do dengue na Paraíba: uma comparação entre os métodos *Scan* flexível e *Scan* circular. Cad Saúde Colet, 2022; 30(4) 561-571. <https://doi.org/10.1590/1414-462X202230040082>

### Resumo

**Introdução:** Estatísticas espaciais são usadas para auxiliar gestores de saúde na tomada de decisão, informando a taxa de ocorrência de agravos na população e destacando quando estas alcançam valores além do esperado. **Objetivo:** Compreender o funcionamento e aplicabilidade das Estatísticas Espaciais *Scan* flexível e *Scan* circular, comparando seus resultados na detecção de aglomerados espaciais usando dados epidemiológicos reais do dengue no estado da Paraíba – Brasil. **Método:** Descreveu-se o processo detalhado da aplicação das estatísticas *Scan* flexível e *Scan* circular para a detecção de áreas significativas de risco (aglomerados) do dengue na Paraíba, nos anos de 2009 a 2013, por meio dos *software FLeXScan* e *SaTScan*. **Resultados:** Ambos os métodos revelaram o oeste do estado como a região com maior frequência de aglomerados detectados com alto risco, em todos os anos analisados, levando-se em consideração os mapas de risco de incidência do dengue na Paraíba, nos anos de 2009 a 2013. **Conclusão:** As estatísticas *Scan* flexível e *Scan* circular são praticamente similares quanto à eficiência na detecção de aglomerados do dengue. Entretanto, verificaram-se problemas de superestimação no método *Scan* circular e subestimação no método *Scan* flexível na detecção dos aglomerados. Destacou-se ainda o auxílio destas estatísticas espaciais aos gestores de saúde quanto à localização das regiões de agravo da doença, tornando mais efetivo o direcionamento das ações de combate de forma politicamente correta. **Palavras-chave:** métodos de aglomeração espacial; estatística *scan* circular; estatística *scan* flexível; epidemiologia do dengue.

### Abstract

**Background:** Spatial statistics are used to help health managers make decisions, informing the rate of occurrence of diseases in the population and highlighting when they reach values beyond expectations. **Objective:** To understand the functioning and applicability of Spatial Statistics Flexible Scan and Circular Scan by comparing their results in the detection of spatial clusters using real epidemiological data of dengue in the state of Paraíba – Brazil. **Method:** The detailed process for applying the flexible scan and circular scan statistics for detecting significant dengue risk areas (clusters) in Paraíba, between 2009 and 2013, was described using the software *FLeXScan* and *SaTScan*. **Results:** Both methods showed



Este é um artigo publicado em acesso aberto (Open Access) sob a licença Creative Commons Attribution, que permite uso, distribuição e reprodução em qualquer meio, sem restrições desde que o trabalho original seja corretamente citado.

Trabalho realizado na Universidade Federal da Paraíba (UFPB) – João Pessoa (PB), Brasil.  
Correspondência: Ana Cláudia Oliveira de Melo. E-mail: [anaclaudiaemelo@gmail.com](mailto:anaclaudiaemelo@gmail.com)  
Fonte de financiamento: CNPq (Processo nº. 308250/2015-0).  
Conflito de interesses: nada a declarar.  
Recebido em: Mar. 11, 2020. Aprovado em: Dez. 26, 2020

the highest frequency of clusters detected at high risk in the western region of the state, in all the years analyzed, considering the risk maps of dengue incidence in Paraíba, between 2009 and 2013. **Conclusion:** The flexible scan and circular scan statistics are practically similar in terms of efficiency in detecting clusters of dengue. However, there were problems of overestimation in the circular Scan method and underestimation in the flexible Scan method in the detection of clusters. It is also worth highlighting that these spatial statistics help health managers locate the regions of disease aggravation, making it more effective to direct combat actions in a politically correct manner.

**Keywords:** spatial agglomeration methods; circular scan statistics; flexible scan statistics; epidemiology of dengue.

## INTRODUÇÃO

A informação do espaço geográfico com taxa de ocorrência de agravos na população acima do esperado promove uma intervenção eficaz na atenção à saúde<sup>1</sup>. É interessante para os gestores, que promovem cuidados em saúde, a tomada de decisão científica baseada no conhecimento das especificidades de cada população e região (geográfica), para determinar as políticas de acordo com os diferentes níveis de prioridade da população assistida<sup>2-4</sup>. Problemas como esses remetem ao uso dos métodos de aglomeração espacial, que se utilizam de informações georreferenciadas para identificar localidades com valores significativos diferenciados ou não do ponto de vista estatístico.

O mosquito *Aedes aegypti*, vetor transmissor do dengue, foi detectado nas principais cidades do Brasil na década de 1970, após ter sido erradicado na década de 1950<sup>5</sup>. O combate à doença é mais efetivo quando é possível detectar uma alta incidência no número de casos registrados. Na epidemiologia, isso ocorre a partir dos registros desses casos, possibilitando a detecção de aglomerados espaciais da doença na região geográfica em estudo. Um aglomerado espacial é definido como um conjunto de áreas, da região (universo) do estudo, que apresentam um risco significativamente alto de um determinado evento, em relação à região considerada, embora não seja esperada essa observação nessas áreas<sup>6</sup>.

Essa detecção serve como base para direcionar as políticas efetivas no combate à doença, assim é importante buscarmos a melhor metodologia disponível. Métodos para detecção de aglomerados espaciais estão disponíveis na literatura científica<sup>7-11</sup>. Cada método produz diferentes resultados, além disso, não há informação de referência para avaliar quais aglomerados são verdadeiros ou não. Assim, são usadas formas indiretas de avaliação, baseadas, por exemplo, nos mapas de risco de incidências<sup>12</sup>.

Este artigo objetiva compreender o funcionamento e aplicabilidade das estatísticas *Scan* circular e *Scan* flexível, comparando seus resultados na detecção de aglomerados espaciais, usando dados epidemiológicos reais do dengue no estado da Paraíba – Brasil. Essas estatísticas são capazes de auxiliar, de forma efetiva, os gestores da saúde no direcionamento dos recursos e políticas de ação às regiões sob risco. Por isso, um tutorial de como aplicá-las pode proporcionar o uso mais frequente destas estatísticas.

## METODOLOGIA

As estatísticas *Scan* circular e *Scan* flexível pretendem detectar aglomerados espaciais (conjunto de sub-regiões) com riscos significativos, quando comparados com o risco da região em estudo, por meio de janelas. Essas janelas assumem formato circular, no caso da estatística *Scan* circular, e no formato flexível (não circular), no caso da estatística *Scan* flexível.

### A estatística *Scan* circular

Considerando a situação em que a região geográfica em estudo (por exemplo, o estado da Paraíba) é dividida em  $m$  sub-regiões ou geo-objetos (municípios, distritos, bairros etc.), a variável aleatória  $N_i$  será o número de casos na sub-região  $i$ , com valor observado  $n_i$  ( $i = 1, \dots, m$ ), e o número total de casos observados será dado então por  $n = n_1 + \dots + n_m$ . Os  $N_i$  são variáveis aleatórias independentes com distribuição de Poisson (este modelo probabilista requer a contagem de casos, para cada sub-região de localização dos eventos), então

$N_i \sim \text{Poisson}(\lambda_i)$ ,  $i = 1, \dots, m$ , onde  $\lambda_i$  denota a média da distribuição de Poisson e é o número esperado de casos da sub-região  $i$ . Assim calculamos  $\lambda_i$  como (Equação 1)<sup>11</sup>:

$$\lambda_i = n \frac{P_i}{\sum_{k=1}^m P_k}, \quad i = 1, \dots, m, \quad (1)$$

onde  $P_i$  denota o tamanho da população na sub-região  $i$ .

A hipótese  $H_0$  afirma que não há aglomerados espaciais na sub-região  $i$ , ou seja, o número de casos notificados dentro da sub-região  $i$  observada é igual ao número de casos no exterior dela, como segue (Equação 2)<sup>6</sup>:

$$H_0 : E(N_i) = \lambda_i, N_i \sim \text{Poisson}(\lambda_i), i = 1, \dots, m. \quad (2)$$

Usaremos as coordenadas do centro do município para especificar a posição geográfica de cada sub-região  $i$ <sup>13</sup>, que chamaremos de centroide. Kulldorff<sup>10</sup> propõe, para a situação descrita acima, a estatística *Scan* circular, que gera uma janela  $Z$  em cada centroide das sub-regiões. Neste caso, uma janela consiste em um círculo criado a partir do centroide. Para qualquer destes centroides, o raio do círculo varia continuamente desde zero até um percentual da população em risco a ser coberta. Esse percentual é preestabelecido pelo pesquisador. Logo, se uma janela contém o centroide de uma sub-região, o raio do círculo crescerá até englobar o percentual da população estabelecido envolvendo outras sub-regiões vizinhas. Definimos  $K$  como o número máximo de sub-regiões dentro da janela, ele será limitado pelo percentual da população em risco. Então denotamos  $Z_{ik}$  ( $k = 1, \dots, K_i$ ) como a janela composta pelos  $k - 1$  vizinhos à sub-região  $i$ , assim todas as janelas a serem verificadas pela estatística *Scan* circular estão incluídas no conjunto  $Z_1$  de todas as janelas circulares, ou seja,  $Z_1 = \{Z_{ik} | 1 \leq i \leq m, 1 \leq k \leq K_i\}$ .

Para verificar se uma janela será ou não considerada como um aglomerado, deveremos realizar um teste de hipóteses para verificar se o número de casos observados dentro da janela é maior que o número de casos esperado no exterior dela. Com a utilização da notação da janela  $Z \in Z_1$ , a hipótese nula (1) é expressa como (Equação 3):

$$H_0 : E(N(Z)) = \lambda(Z), \text{ para todo } Z \in Z_1, \quad (3)$$

onde  $N(Z)$  e  $\lambda(Z)$  denotam, respectivamente, a variável aleatória para o número de casos observados e o número esperado de casos, sob  $H_0$ , dentro da janela especificada. A hipótese alternativa  $H_1$  afirma existir pelo menos uma janela  $Z \in Z_1$ , na qual o risco é mais elevado no seu interior, quando comparado com seu exterior (Equação 4):

$$H_1 : E(N(Z)) > \lambda(Z), \text{ para algum } Z \in Z_1. \quad (4)$$

É possível calcular a verossimilhança do número de casos observados dentro e fora da janela respectivamente, para cada janela de  $Z$ . Kulldorff<sup>10</sup> propõe que sob  $H_0$ , a estatística da razão de verossimilhança seja calculada por (Equação 5):

$$\lambda_K = \max_{Z \in Z_1} \lambda_K(Z). \quad (5)$$

A razão de verossimilhança é uma função que maximiza a informação dos dados, sendo assim,  $\lambda_K$  será o valor esperado que contém o maior teor de informação nas  $K$  janelas observadas.

Kulldorff<sup>14</sup> propôs o *software SaTScan* para realizar o agrupamento das sub-regiões  $i$  em aglomerados de forma circular. No *SaTScan*, é necessário informarmos a matriz de coordenadas dos municípios ( $m$  sub-regiões), tamanhos das populações em cada município ( $P_i$ ), número de casos notificados em cada município ( $n_i$ ). A restrição está relacionada ao percentual da população em risco denominada  $\beta\%$ . A literatura recomenda que este percentual seja menor

que 50% da população em risco, devendo-se analisar diferentes percentuais e escolher aquele que melhor se aproxima do mapa de risco<sup>15,16</sup>.

### A estatística *Scan* flexível

Como a estatística espacial *Scan* circular usa uma “janela circular” para definir a área do aglomerado, é difícil detectar corretamente as áreas não circulares, tais como aquelas ao longo das margens de um rio. A proposta de Tango e Takahashi<sup>11</sup> é criar uma janela  $Z$  de forma flexível (não circular) em cada centroide da sub-região, ligando as sub-regiões vizinhas. O processo ocorre da seguinte forma: para qualquer sub-região  $i$ , criamos o conjunto de janelas de forma flexível com comprimento  $k$ , o que consiste em  $k$  sub-regiões conectadas incluindo  $i$ , e vamos mover  $k$  de 1 até o comprimento máximo preestabelecido  $K$  de vizinhos mais próximos. As sub-regiões ligadas são restritas aos subconjuntos do conjunto de sub-regiões  $i$ , e os  $K$  vizinhos mais próximos à região  $i$ . Ao final, várias janelas diferentes de formas arbitrárias e sobrepostas, umas às outras, são criadas. Seja  $Z_{ik(j)}$ ,  $j = 1, \dots, j_{ik}$  denotando a janela de ordem  $j$ , a qual é um conjunto de  $k$  sub-regiões conectadas a partir da sub-região  $i$ , onde  $j_{ik}$  é a janela  $j$  satisfazendo  $Z_{iK(j)} \subseteq Z_{iK}$  para  $k = 1, \dots, K_i = K$ . Então, todas as janelas percorridas são incluídas no conjunto  $Z_2 = \{Z_{ik(j)} \mid 1 \leq i \leq m, 1 \leq k \leq K_i, 1 \leq j \leq j_{ik}\}$ .

Comparando  $Z_1$  e  $Z_2$ , para qualquer sub-região  $i$ , a estatística *Scan* circular considera  $K$  (limitado por  $\beta\%$ ) círculos concêntricos que denotamos por  $Z_1$ . Já a estatística *Scan* flexível considera  $K$  círculos (cada um centrado em uma sub-região  $i$ ) concêntricos mais todos os conjuntos de sub-regiões ligados (incluindo a única região  $i$ ), cujos centroides estão localizados dentro do  $K$ -ésimo maior círculo concêntrico que denotamos por  $Z_2$ . Portanto, o tamanho de  $Z_2$  tende a ser maior do que o de  $Z_1$ , quando as sub-regiões do conglomerado têm formato não circular. Outro ponto a ser destacado é que, na estatística *Scan* flexível,  $K$  é um parâmetro computacional, não depende da população em risco, seu comprimento máximo deve ser inferior a 30, pois a carga computacional, devido ao grande número de possíveis combinações de janelas, tornar-se-ia muito pesada.

Na estatística *Scan* flexível, se a janela  $Z$  é a janela que contém a razão de máxima verossimilhança, ela contém também o aglomerado mais provável. No entanto, não é interessante que  $Z$  continue aumentando o seu raio, quando já englobou os aglomerados espaciais de maior risco, apenas para atingir o percentual da população preestabelecido pelo usuário, pois assim arriscará englobar também aglomerados espaciais de menor risco<sup>11,17</sup>.

Tango e Takahashi propuseram o *software* *FleXScan*<sup>11</sup> para realizar o agrupamento das sub-regiões  $i$  em aglomerados de forma flexível. No *FleXScan*, precisamos informar a matriz de coordenadas dos municípios (das  $m$  sub-regiões), tamanhos das populações em cada município ( $p_i$ ), número de casos notificados em cada município ( $n_i$ ) e o valor de  $K$  (o valor padrão é 15), que pode variar conforme a necessidade do pesquisador até no máximo 30.

### Mapas da razão de incidência

Para analisar os mapas das estatísticas espaciais, tomamos como referência o mapa da Razão de Incidências Espacial (*RIE*)<sup>18</sup>. Para a obtenção do mapa coroplético, que descreve a incidência do dengue no estado, faz-se necessário o cálculo da *RIE*. Este indicador representa o risco da ocorrência de um fenômeno epidemiológico em uma área  $i$  com relação a toda a região geográfica em estudo e é calculada por (Equação 6)<sup>18</sup>:

$$RIE_i = \frac{x_i / n_i}{\sum_{i=1}^m x_i / \sum_{i=1}^m n_i}, \quad (i = 1, \dots, m), \quad (6)$$

onde  $x_i$  é o número de ocorrência do fenômeno em uma região e  $n_i$  é a população dessa região. As sub-regiões são comparáveis entre si, pois o denominador da *RIE* será o mesmo para todas as regiões. Assim, a incidência da região geográfica em estudo é a mesma no mapa.

## Passos da aplicação do método e definições dos parâmetros

Explorou-se a capacidade de detecção de aglomerados espaciais do dengue na Paraíba (PB) nos anos de 2009 a 2013, segundo as estatísticas espaciais *Scan* circular e *Scan* flexível, acima descritas, percorrendo para isso os passos seguir.

1. As sub-regiões como os  $m$  ( $m = 223$ ) municípios da Paraíba e assim  $i = 1, \dots, 223$ . O número total de casos observados no estado será dado por  $n = n_1 + \dots + n_{223}$ ;
2. A hipótese  $H_0$  afirma que não há aglomerados espaciais na sub-região  $i$  e os  $N_i$  são variáveis independentes de Poisson, logo,  $H_0 : E(N_i) = \lambda_i$ ,  $N_i \sim \text{Poisson}(\lambda_i)$ ,  $i = 1, \dots, 223$ . Calculou-se, assim,  $\lambda_i$  como  $\lambda_i = n \frac{p_i}{\sum_{k=1}^m p_k}$ ,  $i = 1, \dots, 223$ , onde  $p_i$  denota o tamanho da população na sub-região  $i$ ;
3. Para usar os softwares *StatScan* (versão 9.6 (<http://www.satscan.org/>)) e *FleXScan* (versão 3.1.2 (<https://sites.google.com/site/flexscansoftware/>)), informou-se: a matriz de coordenadas dos municípios para definir seus centroides, a matriz com os nomes dos municípios e a matriz com o número de casos registrados e população por municípios. Essas informações servem para o cálculo dos valores esperados, *RIE* e demais medidas necessárias a cada *software*;
4. Na estatística *Scan* circular definiu-se o melhor valor do  $\beta\%$  dentre os seguintes percentuais 0,1%; 0,5%; 0,7% e 1% para controlar o valor do  $K$ ;
5. Para o agrupamento espacial no *software FleXScan*, testaram-se diferentes valores para o  $K$ : 5, 7, 10, 12 e 15, por ano analisado;
6. Para avaliar os resultados de detecção dos aglomerados obtidos pelas estatísticas *Scan* circular e *Scan* flexível, observaram-se os mapas baseados na *RIE* (casos notificados nos anos 2009 a 2013). Destaca-se que, para as duas estatísticas, foram considerados apenas os aglomerados mais significativos (5% de significância).

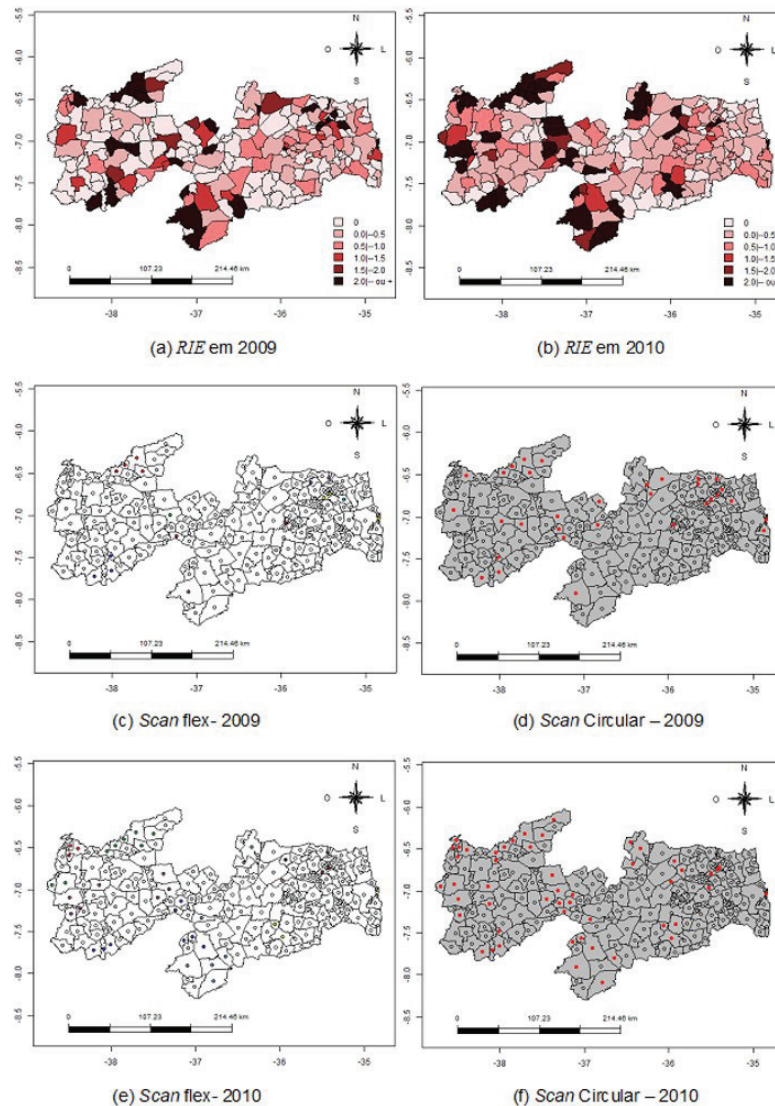
## RESULTADOS

Nesta seção comparamos os resultados dos agrupamentos dos municípios em aglomerados, realizados pelas estatísticas *Scan* circular e *Scan* flexível, com os mapas da *RIE* do dengue nos 223 municípios da Paraíba, nos anos de 2009 a 2013. Para ilustrar o desenvolvimento das análises, apresentamos os referidos mapas da *RIE*, assim como os mapas gerados pelas estatísticas *Scan* flexível e *Scan* circular, respectivamente. De acordo com o ano analisado no *FleXScan*, os melhores valores para  $K$  foram os seguintes: 2009:  $K = 7$ , 2010:  $K = 5$ ; 2011:  $K = 5$ ; 2012:  $K = 5$  e em 2013:  $K = 7$ . No *SaTScan* para todos os anos analisados, obteve-se o melhor resultado para  $\beta\%$  (percentual da população em risco) = 0,01%, ou seja,  $K$  foi controlado por esta proporção.

No mapa da *RIE* classificamos o município considerando a incidência epidemiológica, na Paraíba, da seguinte forma:  $RIE = 0$  indica que não há risco epidemiológico no município, ao comparar-se a incidência no estado (e assim por diante); relativamente baixo ( $0 < RIE < 0,5$ ); baixo ( $0,5 \leq RIE < 1,0$ ); relativamente alto ( $1,0 \leq RIE < 1,5$ ); alto ( $RIE \geq 1,5$  vezes o risco do estado).

A Figura 1a revela que na Paraíba em 2009 há muitos municípios que apresentaram *RIE* igual a 0 (zero). As regiões leste e agreste possuem apenas 11 municípios que registram *RIE* alta, de um total de 32. Os demais municípios com *RIE* elevada encontram-se ao centro e oeste, havendo uma pequena concentração na parte centro-sul do estado. A Figura 1b mostra que houve um aumento na incidência do dengue em 2010, quando comparamos com o mapa da *RIE* em 2009 (Figura 1a). O número de municípios que possuem risco elevado aumentou no centro-sul e oeste do estado. O litoral continua com risco baixo, com exceção do município de Cabedelo, onde o risco permanece alto.

Observamos nas Figuras 1c e 1d, que os dois métodos detectaram os municípios com maior risco exatamente nas regiões oeste, centro e agreste do estado, visto no mapa da *RIE* em 2009. O número de municípios detectados pela estatística *Scan* circular, em 2009, foi maior que na *Scan* flexível.



**Figura 1.** Mapas da *RIE* do dengue na Paraíba, *Scan flex* e *Scan circular* para 2009 e 2010

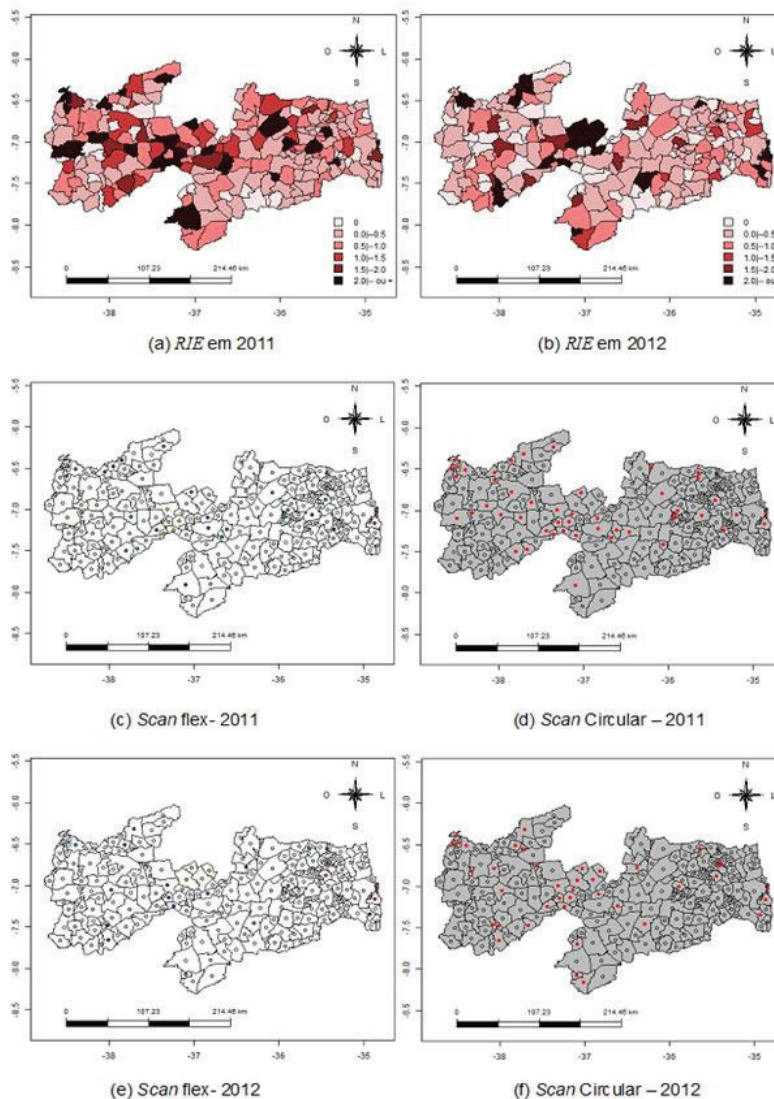
As Figuras 1e e 1f revelam que os dois métodos detectaram os municípios com maior risco de forma aproximada ao mapa da *RIE* em 2010, todavia o *Scan circular* parece detectar alguns municípios com baixo risco; isso não ocorre com a estatística *Scan flexível*.

No mapa da *RIE* em 2011 (Figura 2a), observamos um menor número de municípios com alto risco no oeste do estado e um maior número com risco relativamente alto, em relação ao ano de 2010. Generalizando, podemos dizer que em 2011 o nível do risco, na grande maioria dos municípios, varia de relativamente baixo a relativamente alto.

É visível no mapa da *RIE* em 2012 (Figura 2b) que há menos municípios com alto risco em relação a 2011, assim como uma diminuição de municípios com risco relativamente alto ( $1,0 \leq RIE < 1,5$ ), principalmente no agreste e oeste do estado.

Na Figura 2, verificamos que para 2011 a estatística *Scan flexível* (Figura 2c) apresentou resultado próximo à estatística *Scan circular*, pois ambas detectaram os municípios com alta *RIE* espalhados nas regiões (litoral, agreste, centro e oeste) do estado, como observamos na comparação das duas Figuras 2a, 2c e 2d.

Em 2012 (Figura 2f), o *Scan circular* detectou todos os municípios com alto risco e 7 com risco não elevado, segundo o mapa da *RIE* em 2012 (Figura 2b), enquanto o *Scan flexível* (Figura 2e) detectou apenas 2 municípios com baixo risco.



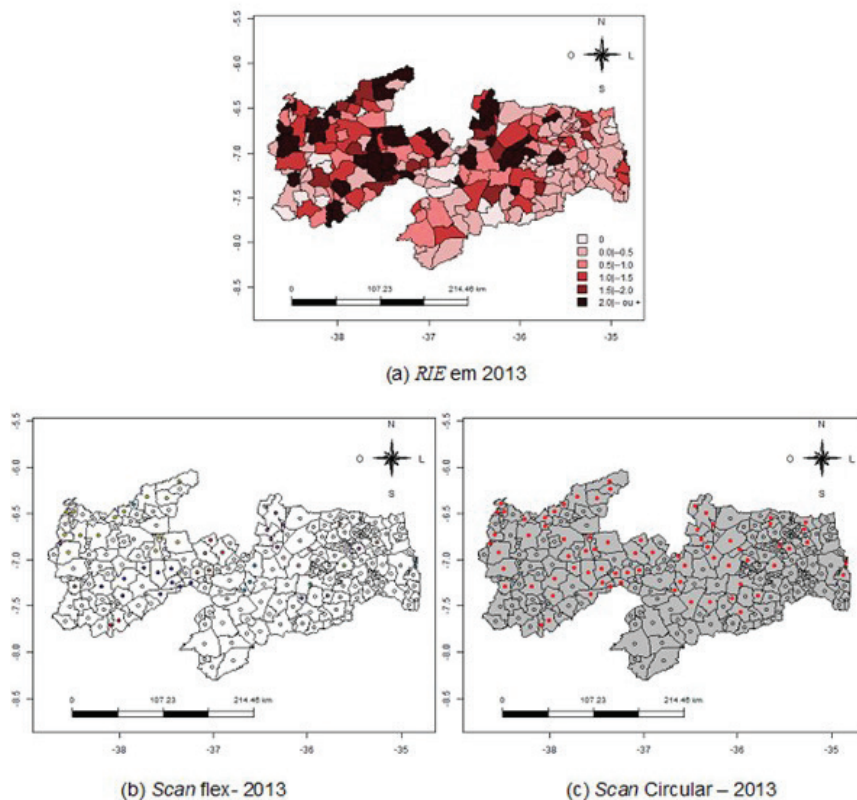
**Figura 2.** Mapas da *RIE* do dengue na Paraíba, *Scan flex* e *Scan circular* para 2011 e 2012

A Figura 3a revela que o maior número de municípios que possuem risco elevado em 2013 encontra-se ao oeste, havendo uma pequena concentração na parte central do estado. As regiões do extremo leste e centro-sul possuem incidência relativamente baixa em comparação com as demais áreas do mapa.

Observamos nas Figura 3b e 3c que, em 2013, os dois métodos detectaram os municípios com alto risco de forma similar. Entretanto, o *Scan circular* detectou 6 municípios com baixo risco, quando o *Scan flexível* detectou apenas um.

Para avaliarmos o desempenho das estatísticas citadas, definimos os agrupamentos das sub-regiões (municípios) que foram incluídas nos aglomerados como concordantes e discordantes, ou seja, consideramos concordantes os agrupamentos dos municípios nos aglomerados com risco alto ( $RIE \geq 1,5$ ) na Paraíba. Consideramos discordantes os agrupamentos dos municípios em aglomerados com  $RIE \leq 1$  (superestimação) ou “o não agrupamento” de municípios nos aglomerados quando  $RIE \geq 1,5$  (subestimação). Para isso, comparamos o mapa dos agrupamentos gerados por cada uma das estatísticas com o mapa da *RIE* do dengue na Paraíba para cada ano analisado.

A Tabela 1 apresenta as seguintes informações (1ª coluna): o item 1 indica o valor da proporção da população em risco, que gerou o melhor agrupamento no *SaTScan*, e o valor



**Figura 3.** Mapas da *RIE* do dengue na Paraíba, *Scan flex* e *Scan circular* para 2013

**Tabela 1.** Resultados das estatísticas *Scan circular* e *Scan flexível*, do dengue na Paraíba nos anos 2009 a 2013

Itens avaliados	Ano									
	2009		2010		2011		2012		2013	
	Circ	Flex	Circ	Flex	Circ	Flex	Circ	Flex	Circ	Flex
1. $\beta\%$ e K	0,01	7	0,01	5	0,01	5	0,01	5	0,01	7
2. Municípios com alto risco no mapa da <i>RIE</i>	32	32	43	43	38	38	32	32	52	52
3. Agrupamentos concordantes	27	20	37	36	38	36	32	26	52	49
4. Subestimações	5	12	6	7	0	2	0	6	0	3
5. Superestimações	3	0	5	1	7	3	7	2	6	1
6. Nº de discordâncias	8	12	11	8	7	5	7	8	6	4
Melhor método	Circular		Flex		Flex		Circular		Flex	

**Fonte:** Dados desta pesquisa

do  $K$  que gerou o melhor resultado no *FlexScan*. O item 2 indica o número de municípios com alto risco ( $RIE \geq 1,5$ ) no mapa da *RIE* da Paraíba. No item 3, temos as concordâncias (nº de municípios agrupados em aglomerados com  $RIE \geq 1,5$ ). Os itens 4 e 5 indicam as discordâncias (nº de subestimações no item 4 e nº de superestimações no item 5). O total de discordâncias, no item 6, resulta da soma de subestimações e superestimações.



Analisando os resultados da Tabela 1, temos que i) Em 2009, considerando o mapa da *RIE* na Paraíba, 32 municípios foram registrados com alto risco, destes, a estatística *Scan* circular detectou 27 com alto risco, subestimou 5 municípios e superestimou 3, cometendo 8 discordâncias ao todo. A estatística *Scan* flexível detectou apenas 20 com alto risco, subestimando 12 municípios e não superestimou, cometendo 12 discordâncias. ii) O mapa da *RIE* em 2010 indica que 43 municípios registraram alto risco. A estatística *Scan* circular detectou 37 municípios com alto risco, subestimou 6 e superestimou 5, cometendo 11 discordâncias. A estatística *Scan* flexível detectou 36 dos 43 municípios com alto risco, subestimou 7 e superestimou 1. iii) Dos 38 municípios com alto risco, indicados no mapa da *RIE* em 2011, a estatística *Scan* circular conseguiu detectar todos, quando comparado ao mapa da *RIE*, porém superestimou 7 municípios e cometeu 7 discordâncias. A estatística *Scan* flexível detectou 36 municípios com alto risco, subestimou 2 e superestimou 3, cometendo 5 discordâncias. iv) Em 2012, o mapa da *RIE* registrou 32 municípios com alto risco. A estatística *Scan* circular detectou 32 municípios, aglomerando todas as áreas de alta incidência, porém cometeu 7 discordâncias ao superestimar 7 municípios. Já a estatística *Scan* flexível detectou 26 destes municípios, subestimou 6 e superestimou 2, cometendo 8 discordâncias. v) Em 2013, o mapa da *RIE* indicou 52 municípios com alto risco, a estatística *Scan* circular apresentou resultado similar à estatística *Scan* flexível, todavia, a *Scan* circular detectou todos os 52 municípios e superestimou 6, cometendo 6 discordâncias, enquanto a estatística *Scan* flexível, que detectou apenas 49 municípios com alta incidência, subestimou 3 e superestimou apenas 1, cometendo 4 discordâncias.

Resumindo, a estatística *Scan* flexível destacou-se em três dos cinco anos avaliados: 2010 (8 discordâncias contra 11 do *Scan* circular), em 2011 (5 discordâncias contra 8 do *Scan* circular) e em 2013 (4 discordâncias contra 6 do *Scan* circular). Observa-se que essas discordâncias no *Scan* flexível foram em sua maioria devidas às subestimações. Já a estatística *Scan* circular destacou-se em dois anos: 2009 (8 discordâncias contra 12 do *Scan* flexível) e 2012 (7 discordâncias contra 8 do *Scan* flexível).

## DISCUSSÃO

Os métodos cometeram concordâncias e discordâncias quanto ao agrupamento de municípios com riscos altos e baixos, quando consideramos o mapa da *RIE* como referência na Paraíba, no período analisado. Os dois métodos subestimaram os aglomerados quando não detectaram municípios com *RIE* alta nos aglomerados.

A estatística *Scan* flexível subestimou, em todos os anos analisados, mais que a estatística *Scan* circular. Isso ocorreu expressivamente no ano de 2009, quando o método *Scan* circular subestimou 5 e o método *Scan* flexível subestimou 12 municípios. Isso deve-se ao formato de construção da janela flexível (não circular). Observamos, ainda, que ambos os métodos superestimaram os aglomerados espaciais, quando incluíram no conglomerado municípios com *RIE* menor ou igual a 1, entretanto, a estatística *Scan* circular superestimou os aglomerados em todos os anos analisados em relação à *Scan* flexível.

A estatística *Scan* flexível destacou-se em três dos cinco anos avaliados (2010, 2011 e 2013), mostrando-se mais eficaz, quando se necessita de resultados mais razoáveis, ou seja, se o desejo é não detectar municípios com baixo risco nos potenciais aglomerados. Já a estatística *Scan* circular destacou-se em dois anos (2009 e 2012), mostrando-se mais eficiente na detecção de municípios de alto risco. Entretanto, pesa sobre ela o problema da superestimação, que inclui no aglomerado municípios com baixo risco, o que não é interessante para os órgãos gestores de saúde, visto que direcionar recursos para um conjunto (aglomerado) formado por municípios de baixo risco é ineficiente.

Estas estatísticas são valiosas, quando necessitamos observar o agravo da doença não exatamente no ponto (no município), mas sim na região (no aglomerado de municípios) de interesse, auxiliando o direcionamento das ações de combate às doenças de forma mais precisa quando tratamos de conjuntos de municípios potencialmente em risco. Estas estatísticas vêm

sendo implementadas e utilizadas sistematicamente no Japão e Estados Unidos para determinar políticas de combate na área de saúde, obtendo êxito efetivo nos resultados<sup>19</sup>.

## CONCLUSÃO

As estatísticas espaciais *Scan* circular e *Scan* flexível ainda são pouco utilizadas na Epidemiologia Espacial no Brasil. Este artigo apresentou um tutorial para o uso dessas estatísticas, visando mostrar o funcionamento dessas ferramentas na detecção de aglomerados espaciais, usando como exemplo dados epidemiológicos reais do dengue na Paraíba no período de 2009 a 2013.

Para a epidemiologia do dengue analisada neste trabalho, pode-se dizer que as estatísticas *Scan* circular e a *Scan* flexível equilibram-se nos seus resultados de forma geral. Existem concordâncias e discordâncias entre os resultados reais e os agrupamentos realizados por essas estatísticas. Deve-se estar atento, ao usar-se essas ferramentas, quanto aos problemas da superestimação e a subestimação dos aglomerados, pois enquanto o *Scan* circular superestima suas detecções, o *Scan* flexível subestima. Essas diferenças certamente se devem ao formato geométrico da janela utilizada por cada método. Sendo assim, talvez o uso dos dois métodos combinados possa levar ao caminho da eficiência com ponderação.

Quanto ao uso dessas estatísticas de detecção espacial, para o auxílio a tomada de decisão dos órgãos gestores da saúde, destaca-se que o ganho não está em observar apenas o risco no município, mas principalmente em direcionar as políticas (recursos, projetos etc.) no combate às doenças para os devidos aglomerados, de acordo com o seu nível de risco na grande região, de forma eficiente.

Como trabalhos futuros, propõe-se estender essa comparação para a epidemiologia de outras doenças, de modo a aprofundar o conhecimento das vantagens e desvantagens de cada método em várias situações distintas.

## AGRADECIMENTOS

Os autores agradecem ao CNPq pelo suporte parcial a esta pesquisa (Processo nº. 308250/2015-0 e bolsa PIBIC/CNPq/UFPB).

## REFERÊNCIAS

1. Lima LMM, Melo ACO, Vianna RPT, Moraes RM. Análise espacial das anomalias congênitas do sistema nervoso. *Cad Saude Colet.* 2019;27(3):257-63. <http://dx.doi.org/10.1590/1414-462x201900030313>.
2. Bailey L, Vardulaki K, Langham J, Chandramohan D. *Introduction to epidemiology*. 1st ed. London: Open University Press; 2007.
3. Sanderson C, Gruen R. *Analytical models for decision making*. London: Open University Press; 2006.
4. Rothman K, Lash T, Greenland S. *Modern epidemiology*. Philadelphia: Wolters Kluwer; 2012.
5. Braga IA, Valle D. *Aedes Aegypti*: histórico do controle no Brasil. *Epidemiol Serv Saude.* 2007;16(2):113-8.
6. Knox EG. Detection of clusters. In: Elliott P, editor. *Methodology of enquiries into disease clustering*. London: Small Area Health Statistics Unit; 1989.
7. Anselin L. *Spatial data analysis with GIS: an introduction to application in the social sciences*. Santa Barbara: National Center for Geographic Information and Analysis; University of California; 1992.
8. Besag J, Newell J. The detection of clusters in rare diseases. *J R Stat Soc Ser A Stat Soc.* 1991;154(1):143-55. <http://dx.doi.org/10.2307/2982708>.
9. Yamada I, Rogerson P, Lee G. GeoSurveillance: a GIS-based system for the detection and monitoring of spatial clusters. *J Geogr Syst.* 2009;11(2):155-73. <http://dx.doi.org/10.1007/s10109-009-0080-1>.
10. Kulldorff M. A spatial scan statistic. *Commun Stat Theory Methods.* 1997;26(6):1481-96. <http://dx.doi.org/10.1080/03610929708831995>.
11. Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *Int J Health Geogr.* 2005;4(11):11. <http://dx.doi.org/10.1186/1476-072X-4-11>. PMID:15904524.

12. Moraes RM, Nogueira JA, Sousa ACA. A new architecture for a spatio-temporal decision support system for epidemiological purposes. In Proceedings of the 11th International FLINS Conference on Decision Making and Soft Computing (FLINS2014); 2014; João Pessoa, Brasil. Hackensack: World Scientific; 2014. p. 17-23. [http://dx.doi.org/10.1142/9789814619998\\_0006](http://dx.doi.org/10.1142/9789814619998_0006).
13. Tango T, Takahashi K. A flexible spatial scan statistic with a restricted likelihood ratio for detecting disease clusters. *Stat Med*. 2012;31(30):4207-18. <http://dx.doi.org/10.1002/sim.5478>. PMID:22807146.
14. Kulldorff M. SaTScan v9.1.1: software for the spatial and space-time scan statistics [Internet]. Information Management Service Inc.; 2011 [citado em 2020 mar 11]. Disponível em: <http://www.satscan.org/>
15. Souza GL Jr, Guedes WC, Mesquita DS. A malária no Amazonas na primeira década do século XXI: uma análise puramente espacial. In: Anais do XIX Simpósio Nacional de Probabilidade e Estatística (SINAPE). São Paulo: Associação Brasileira de Estatística; 2010. CD-ROM.
16. Melo JCS, Melo ACO, Moraes RM. Comparação dos métodos scan circular e flexível na detecção de aglomerados espaciais de dengue. In: Anais da 1ª Escola de Informática Teórica e Métodos Formais (ETMF 2016). Natal: UFRN; 2016. p. 1-9.
17. Tango T. A test for spatial disease clustering adjusted for multiple testing. *Stat Med*. 2000;19(2):191-204. [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(20000130\)19:2<191::AID-SIM281>3.0.CO;2-Q](http://dx.doi.org/10.1002/(SICI)1097-0258(20000130)19:2<191::AID-SIM281>3.0.CO;2-Q). PMID:10641024.
18. Lima LMM, Sá LR, Macambira AFUS, Nogueira JA, Vianna RPT, Moraes RM. A new combination rule for Spatial Decision Support Systems for epidemiology. *Int J Health Geogr*. 2019;18(1):25. <http://dx.doi.org/10.1186/s12942-019-0187-7>. PMID:31706302.
19. Takahashi K, Tango T. A comparison of SaTScan and FleXScan for outbreak detection and monitoring. *Advances in Disease Surveillance*. 2008;5:67-72.