ORIGINAL ARTICLE

# STATE LEVEL VARIATION OF BREAST CANCER CASES IN 2007 AMONG MALAYSIAN WOMEN

*Gary Low Kim Kuan[1], Panayiotis Papapreponis[2], Yong Mun Hin[3]*

[1]*Department of Population Medicine, Faculty of Medicine and Health Sciences, Universiti Tunku Abdul Rahman, Malaysia, 43000, Kajang, Selangor, Malaysia*
[2]*Laboratory of Microbiology and Immunology, Cancer Hospital of Thessaloniki, 57010 Greece*
[3]*Medical Department, Roche (Malaysia) Sdn. Bhd, 55000 Kuala Lumpur, Malaysia*

## ABSTRACT

*Breast cancer is the commonest cancer in women worldwide. This study examined the use of spatial analysis and mapping to visualise the disease distribution. The geographic units used were the states of Malaysia. Breast cancer data was obtained from the National Cancer Registry Report 2007 and the female population data was obtained from the Malaysian Census 2010. A spatial analysis was used to analyse the data by indirect standardisation of the underlying female population of each state. Sarawak has a high standardised incidence ratio (SIR) of 16.81 compared to all other states of the country where the highest SIR was only up to 2.15. However, the age-standardised rate (ASR) does not reflect so. SIR could provide a comprehensive evaluation of the disease for further research and public health intervention.*

***Keywords**: Breast cancer, GIS, spatial analysis, standardised incidence ratio, SIR, SaTScan*

## INTRODUCTION

Breast cancer is the most common cancer in women worldwide with the majority of breast cancer deaths occurring in developing countries[1]. Breast cancer is also the commonest cancer among Malaysian women, accounting for 32.1% of all female cancers and the commonest cancer type (both sexes), accounting for 18.1% of total cancer cases in 2007[2].

A rise in breast cancer incidence would lead to hefty economic burden to the health care system. The cost of treatment especially medication, follow-up of patient and treatment of complication are contributing factors to the health care cost[3,4]. Due to the rise in cases of breast cancer, most research primarily focus on its clinical epidemiological aspect. Numerous risk factors of breast cancer have been identified such as, age, race, exposure to oestrogen, early menarche, late menopause, family history, high fat diet and vitamin D deficiency[5-16]. Few studies have investigated the association between the geographical distribution of breast cancer cases and the geographical distribution of socioeconomic status[13,17-22]. These studies were conducted mostly in developed countries and most of the study agreed that with different geographical variation, the underlying population characteristics, health care facility and socioeconomic status played an important role in predicting the incidence of breast cancer.

Based on the current literature, there has been no study conducted to evaluate the geographical variation of breast cancer in Malaysia. In this study, we examined the use of spatial analysis and mapping with the use of spatial analysis software and geographical information system (GIS) to visualise the disease distribution.

## METHOD

### Study Area, Geographic Units, and Geocoding

The area under study was the entire Malaysia and the geographic units used were all states of the country (i.e. the 13 states plus the Federal Territory of Kuala Lumpur). For the needs of the spatial analysis, the geographic position of each geographic unit has to be specified and this was done by the geographic coordinates (longitude and latitude) of its centroid. The centroid of each state was defined as the central location of its highest populated city (i.e. population centroids).The longitude and latitude of the selected centroids were determined by using Google Earth™ software, version 7. The values of longitude (easting) and latitude (northing) were specified in the format of decimal degrees.

### Breast cancer data

We used routinely collected incidence data at the state level obtained from the National Cancer Registry (NCR) Report 2007. This is the latest report during the time period when this study was conducted. It contains information such as the number of cancer cases reported by clinicians from hospitals in Malaysia, and the age-standardised rate (ASR) of the different states of Malaysia[2]. The ASR is an indicator derived from the method of direct standardisation of rates[23]; it expresses the incidence rate of a disease that a certain population would experience, if this population had the age-structure of a standard population, while retaining its own age-specific rates. The standard population used in the NCR report was the World Health Organisation's world standard population[2].

**Population data**
The 2010 Population and Housing Census of Malaysia (Census 2010) provided data on the female population count for each of the states. This is the latest available population data as census is conducted once every 10 years[24].

**Statistical analysis**
Spatial statistical analysis was performed by employing the spatial scan statistic of the SaTScan software, version 9.3[25]. The spatial scan statistic used by SaTScan imposes a circular window on the centroid of each geographic unit of a study area; the size of the circular window is allowed to increase gradually, thus including a number of adjacent geographic units. This way, the window scans the entire study area by investigating numerous combinations of adjacent geographic units as potential clusters. The population included within each scanning window is compared with the population outside it, in respect to the burden of a disease. The statistical calculations involve a maximum likelihood ratio test, which determines whether a studied window constitutes a cluster, or not. The results output produced by SaTScan comprises information about the geographic extend of an identified cluster (i.e. the number of geographic units included, the coordinates of its centroid, and its radius) and also information about the burden of the disease in the clustered area (i.e. the number of observed and expected cases, standardised incidence ratio (SIR) and the observed incidence rate)[25].

The SIR calculated for each identified cluster is derived from the method of indirect standardisation by considering the entire study area as standard population[25]. The SIR is another indicator derived from the method of indirect standardisation of rates. SIR expresses the excess of observed cases of a disease in a certain population, compared with the expected ones, if this population experiences the disease rates of a standard population. The standard population used for the computation of SIR was the entire population of Malaysia. The crude rate (CR) for each state (table 1) was calculated from our crude data by dividing the number of breast cancer cases for each state by the total female population in each state.

It was assumed that the number of breast cancer cases in each state follows the Poisson distribution and, thus, the statistical model used in SaTScan was the Poisson model. As there were no available data on any socioeconomic or behavioural characteristics of the cases in the NCR report 2007, it was not made possible to include covariates in the analysis for confounding control. An identified cluster was considered statistically significant at the level of 0.05 or less.

**Mapping**
The identified clusters were mapped on thematic maps by using ArcGIS software, version 8[26]. A cluster is depicted as an aggregation of adjacent states on a base map of Malaysia.

**RESULTS**

Table 1 displays the absolute number of cases, ASR, CR and SIR of breast cancer in each of the states of Malaysia. The states of Johor and Kedah presented the highest absolute number of cases with 608, followed by 541 cases in Selangor. The lowest was observed in Perlis with only seven cases.

We used the ASRs provided to account for the age parameter in order to compare them among the states. Pulau Pinang presented the highest ASR (46.7 per 100,000), while similar values were observed for Johor, Kuala Lumpur and Melaka. Figure 1 displays the ASR of breast cancer in Malaysian states in 2007.

Sarawak recorded the highest CR with 482.9 per 100,000, followed by Kedah (63.2), Pulau Pinang (42.7), Johor (38.5), Kuala Lumpur (36.5) and Melaka (33.0).The lowest CR observed is in Perlis (5.9 per 100,000).

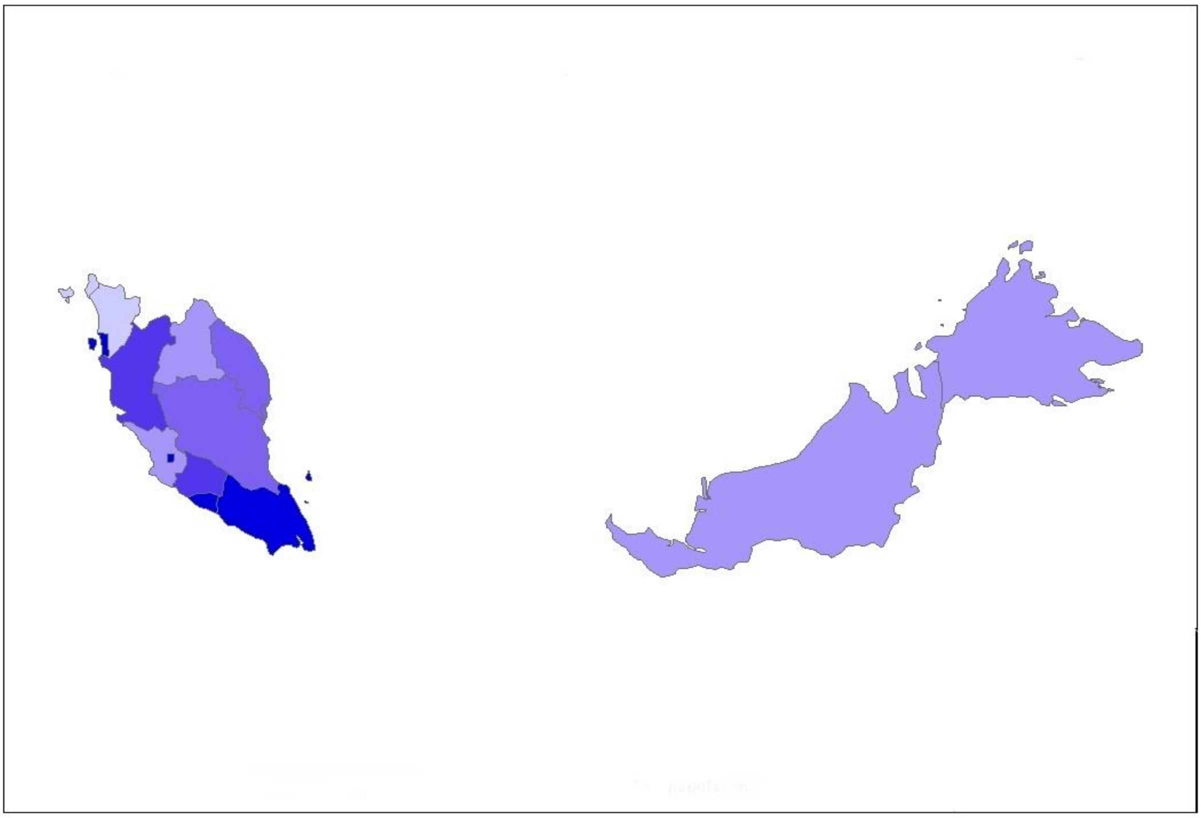
In respect to SIR, the state of Sarawak presented a remarkably high value (16.81) compared to all other states of the country where the highest SIR was only up to 2.15. Seven states (Selangor, Pahang, Kelantan, Terengganu, Sabah, Negeri Sembilan and Perlis) had lower observed cases than expected cases. The observed cases were slightly higher than expected cases (1% to 21% excess) in three states (Kuala Lumpur, Perak, and Melaka). In three states (Pulau Pinang, Johor and Kedah) the excess of the observed cases ranged from 30% to 115%. Figure 2 depicts the SIR of the states in Malaysia in 2007.

The spatial analysis identified four highly statistically significant clusters. The characteristics of these clusters are displayed in Table 2, where they are sorted by descending log likelihood ratio (i.e. the test statistic). The first cluster ($p<0.001$) consisted of the state of Sarawak alone. This cluster included 190 cases of the disease, while the expected number was estimated at 11.30 (SIR=16.81).

**Table1. Absolute number of cases, ASR, CR and SIR of each state in Malaysia**

| States | Number of cases | ASR (per 100,000 population) | CR (per 100,000 population) | SIR |
|---|---|---|---|---|
| Federal Territory of Kuala Lumpur | 292 | 42.6 | 36.5 | 1.21 |
| Selangor | 541 | 14.9 | 20.5 | 0.70 |
| Perak | 346 | 30.3 | 29.7 | 1.01 |
| Pahang | 133 | 22.0 | 18.9 | 0.64 |
| Kelantan | 102 | 17.1 | 13.3 | 0.45 |
| Pulau Pinang | 333 | 46.7 | 42.7 | 1.46 |
| Johor | 608 | 46.5 | 38.5 | 1.31 |
| Kedah | 608 | 4.7 | 63.2 | 2.15 |
| Trengganu | 107 | 25.5 | 21.1 | 0.72 |
| Melaka | 135 | 41.6 | 33.0 | 1.13 |
| Sabah | 147 | 17.0 | 9.5 | 0.32 |
| Sarawak | 190 | 18.8 | 482.9 | 16.81 |
| Negeri Sembilan | 126 | 28.0 | 25.6 | 0.87 |
| Perlis | 7 | 0.6 | 5.9 | 0.20 |

**Figure 1: ASR of breast cancer in Malaysia in 2007.**



Legends: Age standardised rates

- 0.60-0.70
- 4.71-18.80
- 18.81-25.50
- 25.51-30.30
- 30.31-46.70

The second cluster (p<0.001) included two states Kedah and Pulau Pinang. In this cluster, the observed cases were 1.84 times greater than the expected cases. The third cluster (p<0.001) consisted of the states of Johor and Melaka. In this cluster the observed cases were 1.27 times greater than the expected cases. The fourth cluster (p<0.05) consisted of Kuala Lumpur only. The observed cases in this cluster were at 1.21 times greater than the expected cases.

**DISCUSSION**

The six states with the highest ASR were Pulau Pinang, Johor, Kuala Lumpur, Melaka, Perak and Negeri Sembilan. When the spatial scan statistics was analysed, the six states with the highest SIR were Sarawak, Kedah, Pulau Pinang, Johor, Kuala Lumpur and Melaka.

Sarawak appeared to be the top most state in SIR, characterised by a huge SIR (16.81), compared to other states, which have an SIR of less than 2.2. Nevertheless, Sarawak was not seen as one of the top six states when ASR was applied in the comparisons. This could mean that women in Sarawak are possibly exposed to higher risk factors or new risk factors in the development of breast cancer which is not seen in other states. However, SIR in our analysis should be interpreted with caution because the age variable was not included in the analysis.

The ASR is adjusted for the age, population and gender. This allowed fair comparison among the states using the world standard population. However, in terms of resource allocation, the cost of treatment in each individual case counts and it does not matter if the ASR is lower than the other states. Selangor is an example with 541 cases and ASR was 14.9 cases per 100,000 population, but Kedah has 607 cases with a low ASR of 4.7 cases per 100,000 population. The low ASR in Kedah may be due to the different occurrence of breast cancer cases in other age strata especially below the age strata of 45-59 years old (perimenopausal period)[2]. This might lead to false interpretation where funding allocation from the Ministry of Health to each state could be distributed unequally. In research, these cases might be significant to indicate the changing trend of breast cancer to other age strata. While ASR is a good indicator in allowing comparison by standardising the underlying age strata of a population, it does not reflect the true number of cases and this is misleading when healthcare cost is concern. ASR is also unable to detect the changing trend of breast cancer cases among the age groups. Ideally, for a comprehensive comparison among the states in Malaysia, both the size and the age strata of the population of each state should be included in the analysis.

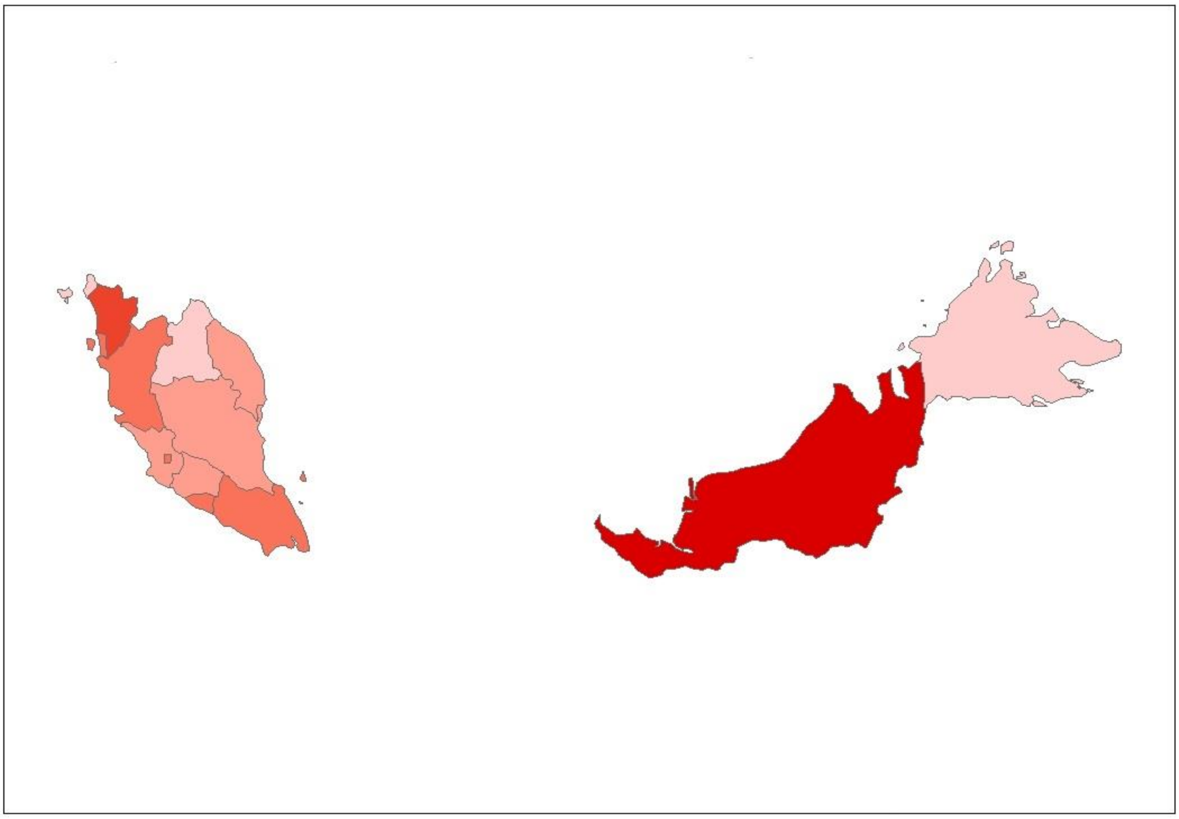
Some studies have used SaTScan for the spatial analysis on breast cancer incidence rate and one of which even conducted the spatial temporal analysis[20,22]. These studies are better in terms of identifying the ‘upstream’ risk factors such as socioeconomic status and urban-rural differences. Better socioeconomic status and living in urban areas were identified as risk factors even in other type of study without standardising the population[13]. Ideally, these ‘upstream’ risk factors and ‘individual level’ risk factors, such as age, exposure to oestrogen, alcohol and smoking should ideally be included in a spatial analysis study. Unfortunately, information on the above variables was not available in the registry. Hence, the ecological fallacy was unavoidable in the present study, since the available data were at group level.

Another limitation of this study is that the data were obtained from the NCR of 2007 which was based on volunteered reporting of cases by doctors. This is evident with the lowest CR reported in Perlis which is due to seven breast cancer cases only. Hence, an incomplete database and under-reporting of cases is expected. However, under-reporting is expected to dilute the true disease burden and give false decreased SIRs. Therefore, the very high SIR observed in Sarawak is a very important finding, taking into account that its true values may be even higher.

**Table 2.Significant spatial clusters**

| Cluster | Central state | States Included | P-value | Observed cases | Expected cases | Incidence Rate per 100,000 | SIR |
|---|---|---|---|---|---|---|---|
| 1 | Sarawak | Sarawak | <0.001 | 190 | 11.30 | 493.3 | 16.81 |
| 2 | Kedah | Kedah, Pulau Pinang | <0.001 | 941 | 510.73 | 54.1 | 1.84 |
| 3 | Johor | Johor, Melaka | <0.001 | 743 | 583.45 | 37.4 | 1.27 |
| 4 | Kuala Lumpur | Kuala Lumpur | <0.05 | 292 | 241.20 | 35.5 | 1.21 |

**Figure 2: SIR of breast cancer in Malaysia in 2007**



Legends: Standardised incidence ratio

| | |
|---|---|
| ▢ | 0.20-0.45 |
| ▢ | 0.46-0.87 |
| ▢ | 0.88-1.46 |
| ▢ | 1.47-2.15 |
| ▢ | 2.16-16.81 |

Geographical units used in the spatial analysis were at the state level is another limitation. These are quite large regions and, thus, spatial precision is low; smaller geographical units are always preferred to allow identification of precise spatial variation of the disease of concern.

A robust study is recommended to overcome these limitations. A cross-sectional study is appropriate to obtain information on risk factors at the individual level and to control for the spatial variation of those factors in the spatial analysis. This will provide better insight as to whether the geographical variation of breast cancer is significant or not after adjusting for these risk factors. This will also provide information on the clustering of cases, if any, so that a possible source of risk factor could be identified geographically.

**CONCLUSION**

Spatial analysis which uses the SIR as the basis of indirect standardization could provide a more comprehensive evaluation of the disease. SIR

may indicate unusually high prevalence of breast cancer cases in a particular region by standardising the population while ASR, though standardised by age, could not reflect the burden of the disease. It is essential especially for the public health policy makers, politicians and researchers to focus on the needs of each region, thereby preventing health inequalities. However, due to the limitations, this study could only generate hypotheses for a more precise analytical study.

**REFERENCES**

1. World Health Organization. Breast cancer: prevention and control, 2014. Available at: http://www.who.int/cancer/detection/b

reastcancer/en/index1.html#. (accessed May 27, 2014)

2. Zainal Ariffin O, Nor Saleha IT. National Cancer Registry Report 2007. Malaysia Cancer Statistics-Data and Figure, 2011.

3. Arozullah AM, Calhoun EA, Wolf M, et al. The financial burden of cancer: estimates from a study of insured women with breast cancer. *J Support Oncol*. 2004;**2**(3):271-278.

4. Shih YT, Xu Y, Cormier JN, et al. Incidence, treatment costs, and complications of lymphedema after breast cancer among women of working age: a 2-year follow-up study. *J Clin Oncol*. 2009;**27**(12):2007-2014.

5. Pike MC, Krailo MD, Henderson BE, Casagrande JT, Hoel DG. "Hormonal"risk factors,"breast tissue age"and the age-incidence of breast cancer. *Nature*. 1983;**303**(5920):767-770.

6. Kelsey JL, Gammon MD, John EM. Reproductive factors and breast cancer. *Epidemiol Rev*. 1993;**15**(1):36.

7. Gorham ED, Garland CF, Garland FC. Acid haze air pollution and breast and colon cancer mortality in 20 Canadian cities. *Can J Public Health*. 1988;**80**(2):96-100.

8. Armstrong B, Doll R. Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. *Int J Cancer*. 1975;**15**(4):617-631.

9. Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast disease. *New Engl J Med*. 1985;**312**(3):146-151.

10. Garland FC, Garland CF, Gorham ED, Young JF. Geographic variation in breast cancer mortality in the United States: a hypothesis involving exposure to solar radiation. *Prev Med*. 1990;**19**(6):614-622.

11. Garland CF, Garland FC, Gorham ED. Calcium and vitamin D: their potential roles in colon and breast cancer prevention. *Ann N Y Acad Sci*. 1999;**889**(1):107-119.

12. Hisham AN, Yip C-H. Overview of breast cancer in Malaysian women: a problem with late diagnosis. *Asian J Surg*. 2004;**27**(2):130-133.

13. Robert SA, Trentham-Dietz A, Hampton JM, McElroy JA, Newcomb PA, Remington PL. Socioeconomic risk factors for breast cancer: distinguishing individual-and community-level effects. *Epidemiology*. 2004;**15**(4):442-450.

14. Yip CH, Taib NA, Mohamed I. Epidemiology of breast cancer in Malaysia. *Asian Pac J Cancer Prev*. 2006;**7**(3):369.

15. McPherson K, Steel C, Dixon JM. ABC of breast diseases: breast cancer—epidemiology, risk factors, and genetics. *Br Med J*. 2000;**321**(7261):624.

16. Yin D, Morris C, Allen M, Cress R, Bates J, Liu L. Does socioeconomic disparity in cancer incidence vary across racial/ethnic groups? *Cancer Causes Control*. 2010;**21**(10):1721-1730.

17. McLafferty S, Wang F. Rural reversal? *Cancer*. 2009;**115**(12):2755-2764.

18. Mullee MA, De Stavola B, Romanengo M, Coleman MP. Geographical variation in breast cancer survival rates for women diagnosed in England between 1992 and 1994. *Br J Cancer*. 2004;**90**(11):2153-2156.

19. Fukuda Y, Umezaki M, Nakamura K, Takano T. Variations in societal characteristics of spatial disease clusters: examples of colon, lung and breast cancer in Japan. *Int J Health Geogr*. 2005;**4**(1):16.

20. Sheehan TJ, DeChello LM, Kulldorff M, Gregorio DI, Gershman S, Mroszczyk M. The geographic distribution of breast cancer incidence in Massachusetts 1988 to 1997, adjusted for covariates. *Int J Health Geogr*. 2004;**3**(1):17.

21. Barry J, Breen N. The importance of place of residence in predicting late-stage diagnosis of breast or cervical cancer. *Health Place*. 2005;**11**(1):15-29.

22. MacKinnon JA, Duncan RC, Huang Y, et al. Detecting an association between socioeconomic status and late stage breast cancer using spatial analysis and area-based measures. *Cancer Epidemiol Biomarkers Prev*. 2007;**16**(4):756-762.

23. Silva I dos S. Cancer Epidemiology: Principles and Methods. Lyon: International Agency for Research on Cancer, 1999.

24. Department of Statistics. Population Distribution by Local Authority Areas and Mukims, 2010.

25. Kulldorff M. A spatial scan statistic. *Commun Stat Theor M*. 1997;**26**:1481-1496.

26. ESRI (Environmental Systems Resource Institute). ArcMap 9.2. California: ESRI; 2009.