USING A GEOGRAPHIC INFORMATION SYSTEM (GIS) IN EVALUATING THE ACCESSIBILITY OF HEALTH FACILITIES FOR BREAST CANCER PATIENTS IN PENANG STATE, MALAYSIA

Narimah Samat*, Dina Jambi and Norul Shahira Musa
School of Humanities
Universiti Sains Malaysia, Pulau Pinang

Aishah Knight Abd. Shatar
Advance Medical and Dental Institute
Universiti Sains Malaysia, Pulau Pinang

Azizah Ab Manan and Yasmin Sulaiman
Jabatan Kesihatan Negeri
KOMTAR Pulau Pinang

*Corresponding author: narimah@usm.my

Geographical Information Systems (GIS) have been used widely in many developed countries to map health-related events, and the results are used for planning of health services (such as locating screening centres) and in assessing clusters of cases to help identify possible aetiological factors. In the United States in particular, cancer cases notified to registries are routinely entered into GIS, and trends are monitored over time. The Penang Cancer Registry (PCR) is a population-based registry that collects and collates data of all cancer cases diagnosed in Penang as well as cancer cases diagnosed elsewhere of Penang residents. Cancer case reports are generated by providing counts of cases based on Penang home addresses. Mapping of cases using information from the PCR gives a fairly complete picture of cancer cases from the Penang state, and any clusters of cases can be readily identified. This study demonstrates the application of GIS in the mapping and evaluation of the accessibility to health facilities of breast cancer cases in the Penang state. The ArcGIS 9.3 software was used to map and evaluate the spatial clustering of cancer cases, and the Network Analysis function was used to determine the distance between breast cancer cases and health facilities. Although the Penang state is considered one of the more...
developed states in Malaysia with good health facilities, in some parts of the state, health facilities are quite inaccessible. The study results suggest that new health facilities with screening and treatment services should be built in the south of the Seberang Perai and Balik Pulau areas. In addition, this study provides the opportunity to include geographical factors in examining cancer data, which gives a fresh outlook on the issue.

Keywords: GIS, breast cancer, point pattern analysis, network analysis

INTRODUCTION

Cancer is a major health problem in many countries. It has been estimated that there are approximately 24.6 million people living with cancer worldwide in 2002. In the same year, 10.9 million new cancer cases were reported, and the incidence is estimated to continue to increase by approximately 50% by the year 2020 (American Cancer Society, 2003). Cancer is a major concern for both developed and developing countries. Table 1 shows the number of new cancer cases and deaths in males and females by world geographical area. This table reveals that cancer is one of the major causes of death in most parts of the world. In Peninsular Malaysia, for example, a total of 21,773 new cancer cases were diagnosed and registered with the National Cancer Registry among Malaysians in the year 2006. It was comprised of 9,974 males and 11,799 females. The five most common cancers in the Peninsular Malaysian population in 2006 were breast, colorectal, lung, cervical and nasopharynx (Zainal Ariffin, Zainudin and Nor Saleha, 2006).

Breast cancer is the most common cancer among women worldwide and the most important cancer regardless of sex. It is estimated that 1.3 million new cases of invasive breast cancer were expected to occur among women in 2007 (Garcia et al., 2007). In 2004, this type of cancer caused 519,000 deaths worldwide (7% of cancer deaths; almost 1% of all deaths) (Wikipedia, 2009). In Malaysia, breast cancer is the most common cancer among women. There were 3,525 female breast cancer cases registered in National Cancer Registry in 2006, which accounted for 16.5% of all cancer cases registered (Zainal Ariffin et al., 2006) If
not detected and treated promptly, breast cancer can metastasise to the lymph glands and other parts of the body, including lungs, bones and liver (MAKNA, 2006). According to the National Cancer Society, 80% of cancer cases are curable if detected early. However, most cancer cases in Malaysia are detected at the late stage, which make them difficult to cure. This can be due to the lack of awareness among the population regarding cancer screening and the inaccessibility of health facilities that are concentrated within urban centres (Lyons, 2004). Hospital and screening facilities are inaccessible to the poor, in particular, who live in areas that are far away from the urban areas and where the public transportation can be quite inefficient (Philips, 1990; Murad, 2007).

Ideally, the distance to a health facility with treatment and screening facilities should be less than 12 kilometres or a 50-minute drive using private transport at a normal speed (Jordon et al., 2004). Accessibility to a hospital is essential to ensure that patients can get the necessary treatment easily.

GEOGRAPHIC INFORMATION SYSTEM AND HEALTH

Breast cancer screening allows for the detection of cancer in otherwise healthy individuals. Mammography and clinical breast examination are the most common screening methods (Garcia et al., 2007). At present, the National Cancer Registry keeps a record of cancer cases in Malaysia. These records must be presented in tabular form for providing the statistical data on the cases. These data lack the spatial component, which could be used not only for keeping track of new cases but also for showing the geographic location, distribution and pattern of cases (Moore and Carpenter, 1999; Rushton et al., 2006). Furthermore, it could also be utilised to relate the incidence with other geographical and environmental parameters (Wang et al., 2008; Levine et al., 2009). Additionally, visualising the distribution and evaluating the accessibility to health facilities might be useful for health practitioners to evaluate the suitability of the location of existing health facilities or to plan for upgrading or building new facilities at more appropriate locations (Murad, 2007). Therefore, in order to incorporate spatial components into the existing tabular data maintained by the National Cancer Registry or other organisations, a global positioning system (GPS) can be used to
Table 1: Estimated number of new cancer cases and deaths by World Area, 2007*  

<table>
<thead>
<tr>
<th>World Area</th>
<th>Cases</th>
<th></th>
<th></th>
<th>Deaths</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Overall</td>
<td>Male</td>
<td>Female</td>
<td>Overall</td>
</tr>
<tr>
<td>Eastern Africa</td>
<td>138,879</td>
<td>151,173</td>
<td>290,052</td>
<td>112,949</td>
<td>116,105</td>
<td>229,054</td>
</tr>
<tr>
<td>Middle Africa</td>
<td>44,022</td>
<td>43,737</td>
<td>87,759</td>
<td>36,921</td>
<td>35,300</td>
<td>72,221</td>
</tr>
<tr>
<td>Northern Africa</td>
<td>71,569</td>
<td>70,500</td>
<td>142,069</td>
<td>58,940</td>
<td>53,261</td>
<td>112,201</td>
</tr>
<tr>
<td>Southern Africa</td>
<td>39,074</td>
<td>39,011</td>
<td>78,085</td>
<td>29,018</td>
<td>25,226</td>
<td>54,244</td>
</tr>
<tr>
<td>Western Africa</td>
<td>73,059</td>
<td>93,235</td>
<td>166,294</td>
<td>58,046</td>
<td>69,992</td>
<td>128,038</td>
</tr>
<tr>
<td>Eastern Asia</td>
<td>2,003,760</td>
<td>1,309,862</td>
<td>3,313,622</td>
<td>1479,263</td>
<td>847,252</td>
<td>2,326,515</td>
</tr>
<tr>
<td>South-Central Asia</td>
<td>687,977</td>
<td>765,746</td>
<td>1,451,723</td>
<td>499,186</td>
<td>475,437</td>
<td>974,623</td>
</tr>
<tr>
<td>South-Eastern Asia</td>
<td>299,109</td>
<td>319,714</td>
<td>618,823</td>
<td>232,325</td>
<td>198,323</td>
<td>430,648</td>
</tr>
<tr>
<td>Western Asia</td>
<td>116,706</td>
<td>109,182</td>
<td>255,888</td>
<td>82,701</td>
<td>62,843</td>
<td>145,544</td>
</tr>
<tr>
<td>Caribbean</td>
<td>37,840</td>
<td>35,695</td>
<td>73,535</td>
<td>26,513</td>
<td>21,693</td>
<td>48,206</td>
</tr>
<tr>
<td>Central America</td>
<td>83,348</td>
<td>101,458</td>
<td>184,806</td>
<td>53,010</td>
<td>58,213</td>
<td>111,223</td>
</tr>
<tr>
<td>North America</td>
<td>934,509</td>
<td>810,866</td>
<td>1,745,375</td>
<td>370,052</td>
<td>329,978</td>
<td>700,030</td>
</tr>
<tr>
<td>South America</td>
<td>357,374</td>
<td>375,708</td>
<td>733,082</td>
<td>215,808</td>
<td>201,334</td>
<td>417,142</td>
</tr>
<tr>
<td>Eastern Europe</td>
<td>485,390</td>
<td>454,120</td>
<td>939,510</td>
<td>375,331</td>
<td>288,130</td>
<td>663,461</td>
</tr>
<tr>
<td>Northern Europe</td>
<td>227,867</td>
<td>220,844</td>
<td>448,711</td>
<td>134,983</td>
<td>118,724</td>
<td>253,707</td>
</tr>
<tr>
<td>Southern Europe</td>
<td>383,958</td>
<td>291,006</td>
<td>674,964</td>
<td>229,937</td>
<td>152,970</td>
<td>382,907</td>
</tr>
<tr>
<td>Western Europe</td>
<td>521,567</td>
<td>428,913</td>
<td>950,480</td>
<td>295,197</td>
<td>226,106</td>
<td>521,303</td>
</tr>
<tr>
<td>Australia/New Zealand</td>
<td>64,068</td>
<td>53,657</td>
<td>117,725</td>
<td>28,458</td>
<td>22,126</td>
<td>50,584</td>
</tr>
<tr>
<td>Melanesia</td>
<td>3,425</td>
<td>4,242</td>
<td>7,667</td>
<td>2,439</td>
<td>2,604</td>
<td>5,043</td>
</tr>
<tr>
<td>Micronesia</td>
<td>319</td>
<td>353</td>
<td>672</td>
<td>239</td>
<td>213</td>
<td>452</td>
</tr>
<tr>
<td>Polynesia</td>
<td>455</td>
<td>460</td>
<td>915</td>
<td>337</td>
<td>278</td>
<td>615</td>
</tr>
</tbody>
</table>

* Excludes nonmelanoma skin cancer

Estimates were produced by applying age-specific cancer rates of a defined geographic region from GLOBOCAN 2002 to the corresponding age-specific population for the year 2007 from the United Nations (UN) population projections (2004 revision). Therefore, estimates for world areas do not sum to worldwide estimates.

Source: Gracia et al. (2007)
record the geographic location of patients based on their addresses (Wang et al., 2008). This data can then be mapped using GIS software to visualise the distribution, evaluate the pattern of cases and identify possible clusters. Such an approach allows for the generation of maps that show the incidence in different areas and for the ability to communicate the information more effectively (Moore and Carpenter, 1999).

In addition to presenting the rate of cases by areas, mapping point locations of cases also allows for the evaluation of the distribution. Point pattern analysis is used to evaluate the physical distribution of point events and test whether there is a significant clustering of points in a particular area (Cromley and McLafferty, 2002). Formerly, point pattern analysis was used in biology or ecology in detecting the distribution patterns of species. Point pattern analysis is an investigation focused on finding patterns in the data, which is composed of points in a spatial region and allows for the measuring of the location of individuals relative to each other (Robinson, 1998). One common application of point pattern analysis is in epidemiology, which is the study of the distribution and determinants of diseases and injuries in populations are studied (Moore and Carpenter, 1999: 143). Diseases such as cholera and dengue can be mapped to help in evaluating significant occurrences of the disease and in assessing the relationship between the diseases and other environmental and spatial characteristics. The study by John Snow, for example, mapped the cases of cholera in relation to water sources in London back in 1854 (as cited from Moore and Carpenter, 1999). That study, which observed an association between cholera and proximity to a contaminated water supply, provided significant contributions to the field of geography and epidemiology.

Point pattern analysis measures whether the cases are systematically organised or structured or the cases are distributed at random (Levine et al., 2009; Moore and Carpenter, 1999). The nearest neighbour index (NNI) is used to analyse the spread of disease. NNI is a calculation of the inter-incident distances, and it helps in assessing the strength of the clustering of point data. The value of NNI ranges from 0 and 2.1491, where the value of 0 means that the distribution is clustered, the value of 1 means the distribution is random and the value of 2.1491 means the distribution is uniform (Robinson, 1998). NNI can easily be calculated
and used in understanding the occurrence of the incidents. It has, however, been criticised for its failure to distinguish between homogenous and random patterns, and it is heavily dependent upon the size of the study area (Moore and Carpenter, 1999).

In addition to evaluating spatial distributions of disease, GIS and spatial analytical capabilities may be helpful in the study of health care and health care delivery. For example, GIS can be used to evaluate the location of existing health facilities and their accessibility, the concomitant utilisation of health care services and the spatial structure of disease patterns in both static and dynamic forms (Moore and Carpenter, 1999). The study by Wang et al. (2008) for example, evaluated the role of access to health care in explaining the variation of late-stage diagnosis of breast cancer in Illinois using GIS and spatial analysis methods. Studies have found that poor geographical access and socio-economic factors contributed to late-stage breast cancer diagnosis (Lyons, 2004; Rushton et al., 2006).

In order to ensure that people in Penang State have access to health facilities, mapping point locations of cases and evaluating the service area of existing facilities is essential to evaluate the suitability of the location of existing facilities. Such analyses would provide essential information for health practitioners to understand the problem and help them plan for better services to serve patients.

METHODOLOGY

The study mapped the distribution of breast cancer cases in Penang State using addresses obtained from the Penang Cancer Registry (PCR). Figure 1 shows the methods used in this study. Approval was obtained from the Research Ethics Committee (Human) of Universiti Sains Malaysia, which was in compliance with the ICH GCP guidelines (USMKK/PPP/JEPM[204.3 (8)]). This study was also registered with the National Medical Research Register (NMRR-08-1336-2734). The addresses of patients, which were generalised to maintain the confidentiality of patients, were made available by the PCR. The addresses of confirmed breast cancer cases recorded from 1994 to 2003 were provided in Microsoft EXCEL format. A field survey was
GIS in Evaluating the Accessibility of Health Facilities

undertaken to record the geographic coordinates of these cases. As the addresses of the patients were generalised to maintain the confidentiality of the patients, it was often difficult to record the precise location. Various approaches were used to identify the location as close as possible to the actual location. These included checking the addresses by using Google Earth software prior to undertaking the fieldwork, asking the postmen in the study area and getting help from the Department of Town and Country Planning, Penang state. After recording these locations using GPS, the data were loaded into the ArcView 3.2 GIS software and converted into ArcGIS 9.3 format.

In addition to the location of breast cancer cases, the geographic coordinates of all health facilities (private hospitals, public hospitals, public health clinics, private health clinics and state health offices) were recorded. The addresses of these facilities were obtained from the Penang Health Department, and their spatial locations were gathered using GPS. These facilities were mapped to evaluate their distributions in relation to breast cancer cases in the Penang state. The data obtained from GPS were also loaded into the ArcView 3.2 GIS software and later converted into the ArcGIS 9.3 format. Other spatial data such as roads, land use, slope and population were also gathered from other government departments in the Penang state. Roads and land use data were obtained from the Department of Town and Country Planning, Malaysia. The slope was derived from the contour map that was originally digitised from the Topographic Map obtained from the Department of Survey and Mapping, Malaysia (1987). The population data for the year 2000 and 2008 were obtained from the Department of Statistics, Malaysia. These data were combined to form a database for this study. All data were in ArcGIS 9.3 format.
Point pattern analysis was conducted to evaluate the distribution of breast cancer cases in the Penang state. NNI was calculated for the evaluation of the pattern of the distribution of the cases. NNI calculates the distance between one point to its nearest points. It was calculated using equation (1) below (Robinson, 1998).

\[
NNI = \frac{d(\text{NN})}{d(\text{ran})}
\]  

(1)

where

- \(NNI\) = ratio of observed nearest neighbour distance to the mean random distance
- \(d(\text{NN})\) = the distance between each point and its nearest neighbour \(d(\text{ran})\) = expected nearest neighbour distance, based on a completely random distribution
GIS in Evaluating the Accessibility of Health Facilities

\[ d(\text{NN}) = \frac{\sum_{i=1}^{N} \text{Min} \left( \frac{d_{ij}}{N} \right)}{N} \]  

(2)

\[ d(\text{ran}) = 0.5 \sqrt[4]{\frac{A}{N}} \]  

(3)

where,

\( \text{Min}(d_{ij}) = \) the distance between each point and its nearest neighbour

\( N = \) the number of points in the distribution

\( A = \) area

This index is used because it gives a systematic measure of the pattern within a specific region. NNI values range from 0 to 2.1491, where the value of 0 indicates that the distribution is clustered, the value of 1 indicates that the distribution is random and the value of 2.1491 indicates that the distribution is perfectly uniform (Robinson, 1998). The Near function was used to calculate the distance from the breast cancer cases to the screening facilities to evaluate the proximity of screening facilities in relation to breast cancer cases. It was calculated based on the distances between one point (location of cases) to another point (centroid of the sub-districts). This analysis was undertaken with the ArcGIS9.3 software.

STUDY AREA AND DATA

The Penang state was chosen as the study area for this study. It is located between 5° 8’ and 5° 35’ in latitude and between 100° 8’ and 100° 32’ in longitude. The study area covers approximately 103,938 hectares and consists of Penang Island and Seberang Perai in the mainland. The state has five administrative districts, namely Seberang Perai Utara, Seberang Tengah, Seberang Perai Selatan, located in the mainland and Timur Laut and Barat Daya districts in the Island (Figure 2). The Penang state is one of the urbanised states in Malaysia where the urban population made up approximately 87% of the 1,313,449 people in 2000.
The Penang state has experienced a significant increase in population; in 1970, the state population was 776,124, and it increased to 900,722 in 1980. The population has continued to increase to 1,064,166 and 1,313,449 in 1991 and 2000, respectively (Department of Statistics, 1996, 2000). In the year 2005, the Penang state population was 1,468,800, is expected to increase to 1,773,442 by 2010 (3.05% annual population growth) and will continue to increase to 2,357,982 by the year 2020 with an estimated 2.89% annual growth rate (Department of Statistics, 2000; Town and Country Planning Department, 2007). By the year 2010, 60% of its population will reside in Penang Island and 40% in Seberang Perai. This structure is expected to change in the year 2020, where 40% of the state population will live in the Island and 60% will reside in the mainland. Health facilities in the Penang state are quite good as there are 6 public hospitals and 11 private hospitals.
The main reason for choosing the Penang State for the study area was due to data availability. The Penang Cancer Registry (PCR) keeps records of the cancer cases whose home addresses are in the Penang State since 1994. This complete dataset is useful in the mapping and evaluation of the distribution of breast cancer cases in Penang. In addition, the Geography Department of the Universiti Sains Malaysia has a collection of databases for the Penang State, which could be used in the analysis. Database building can be costly and time consuming (Longley et al., 2005). Therefore, existing digital data is used wherever possible to reduce the cost of database building. Finally, the researchers were quite familiar with the study area, which made it much easier to locate addresses obtained from PCR.

RESULTS AND DISCUSSION

There were 2,062 breast cases registered with PCR between 1994 and 2003. However, only 1,425 cases (69.1%) could be found during the fieldwork, and the locations were recorded. This was due to the addresses being too general or incomplete. Although various efforts were taken to find the addresses (e.g., asking the personnel in the post-offices, getting help from local people and asking help from the Town and County Planning Department, Penang State), for about 30.9% of the cases, the addresses could not be determined. Figure 3 shows the distribution of breast cancer cases in the Penang State that were located. Most of the uncollected data was in the Timur Laut District. This figure shows that the cases were distributed throughout the state with a high concentration in the Georgetown area. This was probably due to the high population concentration in the Timur Laut District, and more people were screened in this area compared to other parts of the state. Point pattern analysis was undertaken to evaluate the occurrence of clusters. The NNI obtained was 0.3, which showed that the distribution was clustered (Figure 4). In addition, separate NNIs were also calculated for Seberang Perai and Penang Island where the NNI obtained were 0.36 and 0.38, respectively. Significant clusters were found around Georgetown, Bayan Baru, Gelugor and Tanjung Tokong in Penang Island. In the Seberang Perai region, there were clusters around Butterworth and Bukit Mertajam. Many health facilities offering breast cancer screening services were available in municipal areas. Therefore,
it is possible that more cases were detected in urban areas compared to the rural areas due to the greater availability of screening facilities. Further analysis needs to be undertaken to compare the incidence and population density in each sub-district.

Figure 3: The distribution of breast cancer cases in the Penang state
Figure 4: Results of point pattern analysis

The distribution of breast cancer cases was also mapped to show the number of breast cancer cases in each sub-district (Figure 5a). More cases were detected in the Timur Laut district, particularly in the Georgetown and Jelutong areas. This may be due to the location of the screening and treatment facilities in Penang Island. Interestingly, the percentage of cases compared to the population as of the year 2000 revealed a different pattern. More sub-districts in the Seberang Perai region had a higher percentage of cases compared to those in Penang Island (Figure 5b). This study, however, could not identify any specific reasons for the change in the pattern, perhaps due to the unavailability of the data for ethnicity, stage at diagnosis or date of diagnosis.

The distribution of cases can be used to evaluate the occurrence of the cases in the state. Further analysis was undertaken to evaluate the proximity of the incident cases to existing health facilities that offered mammography, which is a type of screening method to diagnose early breast cancer. The Near function in the ArcGIS 9.3 software was used to evaluate the proximity of incident cases to the nearest health facility. Figure 6 shows the distance of the breast cancer cases to the nearest screening facilities. The map shows that most of the areas within Penang...
state are near existing health facilities. In Penang Island, for example, Balik Pulau is quite far from any existing screening facility. Although the map shows that most cases were approximately between 8 km and 17 km from existing facilities, the measurements were undertaken "as the crow flies". Thus, it does not incorporate any physical barriers or road networks in the analysis. Further analysis is still being undertaken to include these criteria. In the Seberang Perai region, there are only two facilities offering mammography. The area in the northern part of Seberang Perai Utara (Tasek Gelugor) and the southern part of Seberang Perai Selatan (Nibong Tebal and Sri Ampangan) is quite far from any existing screening facilities (approximately between 16 km and 25 km). Therefore, new facilities are needed to better serve the public in these two areas.
In addition, the study also measured the proximity of each of the sub-districts (mukims) to existing screening facilities. This analysis was undertaken by converting the sub-districts to centroids (centres of sub-districts), and the near function of ArcGIS 9.3 was used to calculate the proximities. The output is presented in Figure 6. This analysis showed that some sub-districts such as those in the Seberang Perai Selatan and Seberang Perai Utara had very poor access, and people had to travel for more than 20 kilometres to get to the nearest screening facility. These
areas, therefore, can be considered as quite inaccessible to a health screening facility using the United Kingdom recommendation of less than 12 km (Jordon et al., 2004).

Figure 7: Distance of screening facilities to sub-districts (mukim)

LIMITATIONS OF THE STUDY

The study mapped confirmed cases that were registered with the PCR. There were 2,062 recorded breast cancer cases between 1994 and 2003 in the Penang State. However, due to the aggregated address data given by the PCR, only 69.1% of the cases could be identified, recorded and mapped. The results were generated after taking this constraint into consideration in evaluating the clustering of the cases. In addition, the study evaluated the distribution of the breast cancer cases in the Penang state, but due to the unavailability of data, the study could not properly explain the spatial distribution of the breast cancer cases in the Penang
State. Only aggregate-level addresses of breast cancer cases were obtained from PCR. Data on ethnicity, stage of cases or number of cases per year were unavailable, which limited the analyses.

Finally, the study measured the proximity of cases to existing screening facilities based on the Near function. This analysis does not take barriers such as heavy traffic or one-way traffic into consideration in evaluating accessibility. It should also be noted that the availability of health facilities in the neighbouring states of Kedah in the north and Perak in the south of the study area had not been taken into consideration. This would affect the accessibility of cases in the mainland but not for cases on the island.

**CONCLUSION**

Breast cancer is the most common cancer among women in Malaysia. This study mapped the distribution of breast cancer cases in the Penang state and found that the cases were clustered in urban areas, with many cases diagnosed in the Timur Laut District. This study, however, was merely observational and the distribution of breast cancer cases might have occurred by chance. The map obtained from this study could not be used to analyse associations between breast cancer incidence and environmental factors. However, it is useful to visualise the distribution and evaluate the accessibility of incidence to existing facilities, which is important for policy makers in planning and providing health services and programs. The study found that more health centres with mammography facilities are needed, especially in the south and the north of Seberang Perai Region and Barat Daya districts in Penang Island.

**ACKNOWLEDGEMENTS**

The authors wish to thank Universiti Sains Malaysia for funding this research under the Research University Grant (1001/PPSK/8120245), the Penang Cancer Registry for giving permission to use data for the study and the Town and Country Planning Department, Penang for providing the spatial data used in the study.
REFERENCES


Narimah Samat et al.


