Exploring Spatial Patterns of Colorectal Cancer in Tehran City, Iran

Kamyar Mansori¹, Alireza Mosavi-Jarrahi², Ali Ganbary Motlagh³, Masoud Solaymani-Dodaran¹, Masoud Salehi⁴, Alireza Delavari⁵, Ali Sanjari Moghadam⁶, Mohsen Asadi-Lari¹*

Abstract

Objectives: Colorectal cancer (CRC) may now be the second most common cancer in the world. The aim of this study was to determine whether clusters of high and low risk of CRC might exist at the neighborhood level in Tehran city.

Methods: In this study, new cases of CRC provided from Cancer Registry Data of the Management Center of Ministry of Health and Medical Education of Iran in the period from March 2008 to March 2011 were analyzed. Raw standardized incidence rates (SIRs) were calculated for CRC in each neighborhood, along with ratios of observed to expected cases. The York and Mollie (BYM) spatial model was used for smoothing of the estimated raw SIRs. To discover clusters of high and low CRC incidence a purely spatial scan statistic was applied.

Results: A total of 2,815 new cases of CRC were identified and after removal of duplicate cases, 2,491 were geocoded to neighborhoods. The locations with higher than expected incidence of CRC were northern and central districts of Tehran city. An observed to expected ratio of 2.57 (p<0.001) was found for districts of 2, 6 and 11, whereas, the lowest ratio of 0.23 (p<0.001) was apparent for northeast and south areas of the city, including district 4.

Conclusions: This study showed that there is a significant spatial variation in patterns of incidence of CRC at the neighborhood level in Tehran city. Identification of such spatial patterns and assessment of underlying risk factors can provide valuable information for policymakers responsible for equitable distribution of healthcare resources.

Keywords: Colorectal cancer- spatial analysis- neighborhood- York and Mollie (BYM) spatial model- Tehran

Asian Pac J Cancer Prev, 19 (4), 1099-1104

Introduction

Cancer is one of the most important causes of mortality and morbidity in the world. It is the third cause of death after cardiovascular diseases and traffic accidents in Iran. Annually about 70,000 new cases of cancer occur in Iran (Gourabi, 2011). Among the types of cancers, colorectal cancer (CRC) is the second most common cancer in the world so that every year about one million new cases of this cancer are diagnosed in the around the world and about half of these cases are diagnosed resulted to death (Besag 1991; Stone et al., 2004).

According to report of International Agency for Research on Cancer (IARC), CRC is the fourth most common cancer after cancers of stomach, prostate and bladder in men Iranian with age-standardized incidence rate (ASIR) and age-standardized mortality rate (ASMR) 7.8 and 3.6 per 100,000; respectively. This type of cancer in women Iranian is the third most common cancer after breast and stomach cancer, with ASIR and ASMR 6.4 and 4.6 per 100,000; respectively. In general, the CRC is the third most common cancer in both sex after breast and stomach cancers with a proportion 7% of incidence rate and 7.6% of mortality rate of total cancers (Ferlay et al., 2015).

In Iran, the incidence of CRC in the elderly in compared to western countries is much lower, but its incidence is higher in the younger generation of the country which can be lead to considerable increase burden of disease in the future (Foroutan et al., 2008). From the other hand, the geographical distribution of CRC is not uniform in the country, so that Iran’s major cities especially Tehran has always been one of the cities with the highest incidence of CRC (Ansari et al., 2006). The different study presented many risk factors (such as poor diet, low physical activity, being overweight, and smoking and alcohol consumption) for CRC incidence (Haggar and Boushey, 2009). Regardless of these risk factors, the various studies have shown that incidence and mortality of CRC can be associated with place and area based risk
Kamyar Mansori et al.

Asian Pacific Journal of Cancer Prevention, Vol 19

Observed cases, the number of expected cases and relative have a Poisson distribution. Raw standardized incidence ratio (SIR) observed in each neighborhood in Iran, the aim of this study was to be dispersed due to extra Poisson variability or over-dispersion. To deal with this problem, the York and Mollie (BYM) spatial model can be used for smoothed the raw SIRs.

BYM model

When using of Poisson model for the count data in the spatial analysis, an important problem is Overdispersion or extra-Poisson variability. It occurs in the presence of spatial autocorrelation in the residual values which lack of attention to this problem can be lead to misleading results in the analysis. For deal with this problem Hierarchical Models such the Besag-York-Mollie (BYM) model is used (Pedigo and Aldrich, 2011). This model assumes that the number of cases of outcome is independent in each neighborhood and follows of the Poisson distribution. The BYM model was introduced by Kaldor and Clayton and was developed by Besag et al., (1991). This model in addition to of independent variables, it considers two sources of changes for the heterogeneity of incidence rate in every neighborhood, that these are \( \nu_i \) and \( u_i \). The formula of BYM model is as follows:

\[
\log (\theta_i) = \alpha + u_i + \nu_i + \sum_{h=1}^{H} \beta_h x_{ih}
\]

In the above formula \( \alpha \) is a log-relative risk baseline, \( \nu_i \) and \( u_i \) are random components of BYM Model and present Non-structural heterogeneity and structural heterogeneity; respectively. For variables of \( \nu_i \) and \( u_i \) are considered prior normal distribution and normal distribution of conditional autocorrelation; respectively. And also for \( \beta \) is considered prior normal distribution. Considering the distribution of conditional autocorrelation, the incidence rate in each neighborhood is dependent to the adjacent neighborhoods and it effected of incidence rate the neighborhoods (Lawson et al., 1999; Lawson et al., 2003).

Materials and Methods

The Study Area and Data collection

This retrospective study conducted in Tehran, Capital of Iran. The Tehran metropolitan area 638 square kilometer is situated on the southern slopes of the Alborz Mountains at a latitude of 35°45′N and a longitude of 51° 25′E. This city consisting of 22 municipal districts and also geographical unit of the study was 374 neighborhoods in Tehran city. In this study, total cases of incident CRC investigated during 2008 to 2011 in Tehran city. The information of these patients with home addresses extracted from Iran’s cancer registry of ministry of health, then geocoded to neighborhood location. Also population of aged 55 and over for each neighborhoods extracted from national census 2006 and 2011.

Statistical analysis

Raw standardized incidence ratio (SIR)

The number of the observed cases in each neighborhood have a Poisson distribution

\[ O_i \sim \text{Poisson}(E_i \theta_i) \]

So that \( O_i \), \( E_i \) and \( \theta_i \) represent the number of the observed cases, the number of expected cases and relative risk for neighborhood \( i \); respectively. Expected cases is calculated as follows:

\[ E_i = n_i \left( \frac{\sum_{j=1}^{n} y_j}{\sum_{j=1}^{n} n_j} \right), \quad i = 1, 2, \ldots, l \]

In the above formula, \( n_i \) is the number of population 50 years and over in neighborhood \( i \). \( y_i \) is the number of the observed cases in the neighborhood \( i \). The standardized incidence ratio (SIR) is equal to the observed to expected ratio. The raw SIRs per neighborhoods were expected to be dispersed due to extra Poisson variability or over-dispersion. To deal with this problem, the York and Mollie (BYM) spatial model can be used for smoothed the raw SIRs.
model by using SaTScan software (v9.4.2) were used for determination of Neighborhood variation in incidence of CRC. For analysis, we need to the number of cases, the number of population and the geographical coordinates (longitude and latitude) by separating each neighborhoods. The standard purely spatial scan statistic creates a circular window (spatial cluster) on the map and then for comparing the number of disease cases in a geographic area \((\text{th}_n)\) with disease cases outside that area \((\text{th}_m)\) it moves throughout the study area. Since the result of the analysis may be sensitive to model parameters, especially window size, Gini coefficient is used for definition of the maximum spatial cluster. Indeed, for determination of the best collection and non-overlapping of clusters the Gini coefficient is more intuitive and systematic way (Han et al., 2016).

For cancer incidence, a Poisson model is typical model. The likelihood ratio statistic (LRS) of Poisson distribution (under test hypothesis; \(H_0: \text{th}_n = \text{th}_m; \text{Ha: } \text{th}_n \neq \text{th}_m\)) for a specific window is proportional to 1:

\[
\left(\frac{c}{E[c]}\right)^c \left(\frac{C-c}{C-E[c]}\right)^{C-c}
\]

In the above formula, \(c\) refers to the observed number of CRC cases within window, \(C\) presents the total number of CRC cases, and \(E[c]\) implies on crude expected number of cases within the window under the null hypothesis, \(C-E[c]\) is expected number of cases outside the window.

The randomization testing or Monte Carlo Hypothesis testing is used for statistical significance level of identified clusters because the exact distribution of LRS is unknown. The large number of random dataset are generated and the LRS value for each of random dataset is computed under null hypothesis. The Monte Carlo p-value of a window is computed as\((R_{\text{best}}+1)/(R+1)\), so that \(R_{\text{best}}\) shows the number of random dataset which its LRS is higher than LRS under real dataset and \(R\) is total number of random dataset. A window would be statistical significance at \(\alpha=0.05\) when it’s LRS is higher than approximately 95% of LRS of random dataset. The windows with highest statistical significance likelihood ratio defined as most likely, secondary and tertiary cluster respectively. For statistical significance of Moran’s Index and spatial clusters, \(P\)-value of < 0.05 using 999 permutations used. Sufficient statistical power was provided by with 999 replication in Monte Carlo simulation. The ArcGIS 10.3 software was used for all cartographic manipulations and displays.

**Results**

In total 2,815 new cases of CRC were identified in during March 2008 to March 2011 after removal of duplicate cases, that 2491 case of these had complete postal address and geocoded to neighborhood location. The minimum and maximum cases of CRC in the neighborhoods were 0 and 57; respectively. Figure 1 shows the number of observed cases of colorectal cancer

| Areas with high rates | Optimal Gini coefficient | MSC | Clustered detected | Involved District | At risk population | Observed cases (O) | Expected cases (E) | Annual cases per 100,000 | O/E | RR** | \(p\)-value |
|-----------------------|--------------------------|-----|-------------------|------------------|-------------------|------------------|------------------|--------------------------|-----|------|-----------|
| Areas with low rates  | 0.354                    | 0.03| Primary           | 2,611            | 38,038            | 152              | 59.05            | 129.6                    | 2.57| 2.68 | <0.001    |
|                       |                          |     | Secondary         | 2,3              | 18,421            | 85               | 28.6             | 149.7                    | 2.97| 3.04 | <0.001    |
|                       |                          |     | Tertiary          | 18,21            | 23,017            | 97               | 35.73            | 136.7                    | 2.71| 2.78 | <0.001    |
| Areas with low rates  |                          |     | Primary           | 4                | 36,175            | 13               | 56.16            | 11.7                     | 0.23| 0.23 | <0.001    |
|                       |                          |     | Secondary         | 15               | 33,053            | 11               | 51.31            | 10.8                     | 0.21| 0.21 | <0.001    |

MSC, maximun size cluster; **, Relative Risk is calculated as the observed divided by the expected within the cluster divided by the observed divided by the expected outside the cluster.
in the neighborhoods of Tehran, as that you have seen the majority of cases of CRC occurred in the northern and central areas of Tehran (Figure 1). In this study, The Moran’s Index was 0.05 with p-value<0.05, as a result, the null hypothesis of zero spatial autocorrelation was rejected for all the neighborhoods.

Figure 2 shows the estimated raw SIR of CRC in neighborhoods of Tehran during 2008-2011. The variation range of estimated raw SIR was from 0 to 24.61. The value of raw SIR was 0 in 65 neighborhoods of Tehran, since there were no cases of colorectal cancer in these neighborhoods. In general, 36% of the neighborhoods had SIR higher than 1 and also 64% had SIR lower than 1 (Figure 2).

Figure 3 shows the estimated standardized incidence ratio (SIR) of colorectal cancer incidence using Besag, York and Mollie (BYM) spatial model in Tehran (2008-2011), as that have seen the results of two methods of raw SIR and BYM show that there is neighborhood inequality in incidence of CRC in Tehran city, so that the neighborhoods with higher than expected incidence of CRC located in districts of northern and central of Tehran city (Figure 2-3).

The geographic pattern of most likely clusters of CRC in neighborhoods of Tehran is depicted in Figure 4. The determined clusters of CRC incidence shows a statistical dispersion (Gini index=0.3540). The clusters with a higher than expected incidence were located in the northern areas, western and central city. The clusters with a lower than expected incidence were seen in the northeast and southeast areas of the city (Figure 4).

The specifications of most likely clusters of CRC in neighborhoods of Tehran can be seen in Table 1. The most likely cluster of higher than expected incidence CRC which was located in the northern areas, western and central city, with observed to expected ratio of 2.57 (p<0.001), including neighborhoods in districts of 2, 6 and 11. This means that the incidence of CRC is 2.57 times higher in this cluster compared to other area of city.

The most likely cluster of lower than expected incidence CRC which was located in northeast and south areas of the city, with observed to expected ratio of 0.23 (p<0.001), including neighborhoods in district of 4. This means that the incidence of CRC is 0.23 lower within in this cluster compared to other area of city (Table 1).

Discussion

Different studies have examined the incidence cancer in Tehran by using pathology or the population based registry, however, this research is the first study of neighborhood based incidence CRC in Tehran city. The aim of this study was determination the smoothed standardized incidence ratio (SIR) and discovery of clusters of high risk and low risk of CRC incidence in the neighborhoods level of Tehran city by BYM spatial model during 2008 to 2011. Determination of clusters of diseases can be help in discovering the risk factors, visualizing patterns of distribution and potential disparities of diseases especially cancers, and finally in equitable allocation screening and early detection programs , as also equitable, palliative and therapeutic services for populations at risk in these clusters (Rassaf et al., 2012).

The results showed that the minimum and maximum cases of CRC in the neighborhoods were 0 and 57; respectively. The variation range of estimated raw SIR was from 0 to 24.61 and also 36% of the neighborhoods had SIR higher than 1. The results of two methods of raw SIR and BYM show that there is neighborhood inequality in incidence of CRC in Tehran city, so that the neighborhoods with higher than expected incidence of CRC located in districts of northern and central of Tehran city. The most likely cluster of higher than expected incidence CRC was located in the northern areas, western and central city, including neighborhoods in districts of 2, 6 and 11. The most likely cluster of lower than expected incidence CRC was located in northeast and south areas of the city, including neighborhoods in district of 4.

The results of this study showed that the most likely cluster of higher than expected incidence CRC was located in the northern areas, western and central city, including
Exploring Spatial Patterns of Colorectal Cancer in Tehran City

DOI:10.22034/APJCP.2018.19.4.1099

Asian Pacific Journal of Cancer Prevention, Vol 19

Exploring Spatial Patterns of Colorectal Cancer in Tehran City

Foroutan M, Rahimi N, Tabatabaeifar M, et al (2008). Clinical features of colorectal cancer in Iran: A 15-year review. Asian Pac J Cancer Prev, 9, 5, 225-7.

Bilancia M, Fedespina A (2009). Geographical clustering of lung cancer in the province of Lecce, Italy: 1992-2001. Int J Epidemiol, 240, 45-59.

Brewer CA (2006). Basic mapping principles for visualizing cancer data using geographical information systems (GIS). Am J Prev Med, 30, 25-36.

Chamanparaa P, Moghimbeigi A, Faradmal J, et al (2015). Exploring the spatial patterns of three prevalent cancer latent risk factors in Iran; Using a shared component model. Int J Epidemiol Res, 2, 68-77.

DeChello LM, Sheehan TJ (2007). Spatial analysis of colorectal cancer incidence and proportion of late-stage in Massachusetts residents: 1995–1998. Int J Health Geogr, 6, 20.

Elferink M, Pukkala E, Klaase J, et al (2012). Spatial variation in stage distribution in colorectal cancer in the Netherlands. Eur J Cancer, 48, 1119-25.

Ferlay J, Soerjomataram I, Dikshit R, et al (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in Globocan 2012. Int J Cancer, 1, 136.

Foroutan M, Rahimi N, Tabatabaeifar M, et al (2008). Clinical features of colorectal cancer in Iran: A 15-year review. J Dig Dis, 9, 225-7.
Gourabi BR (2011). Recognition of geographical diffusion esophagus cancer in southwestern of Caspian Sea. *J Am Sci*, 7, 297-302.

Haggar FA, Boushey RP (2009). Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin Colon Rectal Surg*, 22, 191-7.

Han J, Zhu L, Kulldorff M, et al (2016). Using Gini coefficient to determining optimal cluster reporting sizes for spatial scan statistics. *Int J Health Geogr*, 15, 27.

Kelsall J, Wakefield J (1999). Discussion of 'Bayesian models for spatially correlated disease and exposure data', by Best et al. *Bayesian Anal*, 6, 151.

Keyghobadi N, Lotfi MH, Fallahzadeh H, et al (2013). Nutritional factors related to colorectal cancer in the residents of Yazd City, Iran. *J Health Dev*, 2, 171-81.

Larijani B, Shirzad M, Mohagheghi M, et al (2004). Epidemiologic analysis of the Tehran cancer institute data system registry (TCIDSR). *Asian Pac J Cancer Prev*, 5, 36-9.

Lawson A, Biggeri A, Lessaffre E (1999). Disease mapping and risk assessment for public health.

Lawson AB, Browne WJ, Rodeiro CLV (2003). Disease mapping with WinBUGS and MLwiN, John Wiley and Sons.

Le H, Ziogas A, Lipkin SM, et al (2008). Effects of socioeconomic status and treatment disparities in colorectal cancer survival. *Cancer Epidemiol Biomarkers Prev*, 17, 1950-62.

Mahaki B, Mehrabi Y, Kavousi A, et al (2011). Multivariate disease mapping of seven prevalent cancers in Iran using a shared component model. *Asian Pac J Cancer Prev*, 12, 2353-8.

Mohebbi M, Mahmoodi M, Wolfe R, et al (2008). Geographical spread of gastrointestinal tract cancer incidence in the Caspian Sea region of Iran: spatial analysis of cancer registry data. *BMC Cancer*, 8, 137.

Mosavi-Jarahi A, Moini M, Mohagheghi M-A, et al (2007). Clustering of childhood cancer in the inner city of Tehran metropolitan area: a GIS-based analysis. *Int J Hyg Environ Health*, 210, 113-9.

Moshfeghi K, Mohammad-Beigi A, Hamedi-Sanani D, et al (2011). Evaluation the role of nutritional and individual factors in colorectal cancer. *Zahedan J Res Med Sci*, 13, 12-7.

Navarro A, Muñoz SE, Lantieri MJ, et al (2004). Meat cooking habits and risk of colorectal cancer in Cordoba, Argentina. *Nutrition*, 20, 873-7.

Pedigo A, Aldrich T (2011). Neighborhood disparities in stroke and myocardial infarction mortality: a GIS and spatial scan statistics approach. *BMC Public Health*, 11, 644.

Rassaf MR, Ramezani R, Mehrza M, et al (2012). Inequalities in cancer distribution in Tehran: a disaggregated estimation of 2007 incidence by 22 districts. *Int J Prev Med*, 3, 483.

Shah SA, Neoh HM, Rahim S, et al (2014). Spatial analysis of colorectal cancer cases in Kuala Lumpur. *Asian Pac J Cancer Prev*, 15, 1149-54.

Stone WL, Krishnan K, Campbell SE, et al (2004). Tocopherols and the treatment of colon cancer. *Ann N Y Acad Sci*, 1031, 223-33.