Application of Bayesian Multilevel Space-Time Models to Study Relative Risk of Esophageal Cancer in Iran 2005-2007 at a County Level

Sedigheh Rastaghi1, Tohid Jafari-Koshki2,3, Behzad Mahaki1*

Abstract

Background: Reported age standardized incidence rates for esophageal cancer in Iran are 0.88 and 6.15 for females and males, at fifth and the eighth ranks, respectively, of cancers overall. The present study aimed to map relative risk using more realistic and less problematic methods than common estimators. Materials and Methods: In this ecological investigation, the studied population consisted of all esophageal cancer patients in Iran from 2005 to 2007. The Bayesian multilevel space-time model with three levels of county, province, and time was used to measure the relative risk of esophageal cancer. Analyses were conducted using R package INLA. Results: The total number of registered patients was 7,160. According to the results, the three-level model with adjustment for risk factors of physical activity and smoking had the best fit among all models. The overall temporal trend was significantly increasing. At county level, Ahar, Marand, Salmas, Bojnoord, Saghez, Sarakhs, Shahroud and Torbatejam had the highest relative risks. Physical activity was found to have significant direct association with risk of developing esophageal cancer. Conclusions: Given to great variation across geographical areas, many different factors affect the incidence of esophageal cancer. Conducting further studies at the individual level in areas with high incidence could provide more detailed information on risk factors of esophageal cancer.

Keywords: Disease mapping - esophageal cancer - Bayesian multilevel space-time model - Iran

Asian Pac J Cancer Prev, 16 (14), 5787-5792

Introduction

The causes of human death have been shifted from infectious diseases to non-communicable ones; In developed countries, the incidence of cancers is outpacing cardiovascular diseases (Mohammadpour Tahamian et al., 2013). Cancer is one of the main causes of death and disabling factors all around the world. According to the World Health Organization, up to 2020, the incidence of cancer will be increased 50 percent. In fact, cancer is expected to be the first cause of death in 2030 (Goya, 2007; Zarnegar Nia et al., 2011). Nowadays, cancer is the second cause of death in the world (Scholefield, 2000) and Over 70 percent of deaths occur due to cancer in developing and developed countries (Mathers and Boschi-Pinto, 2000). It is the third cause of death in Iran after cardiovascular diseases and accidents. Hence, it makes necessary to search the treatments and preventive methods as the priorities for health officials and medical community (Naghavi, 2006).

Gastrointestinal tract cancers are prevalent, especially in developing countries (Parkin, 1998). According to the report of International Agency For Research on Cancer, esophageal cancer is the eighth prevalent cancer and sixth cause of death induced by cancer around the world (Mathers and Boschi-Pinto, 2000; Parkin et al., 2005; Semnani et al., 2005; Kamangar et al., 2006; Ma and Yu, 2006; Ferlay et al., 2010; Scarpa et al., 2011). Esophageal cancer is rare in western countries as around 80 percent of all total esophageal cancer cases are from developing countries (Parkin et al., 2005). The prevalence of esophageal cancer is about 5-10 per 100,000 population in the north of America and western Europe (Kamangar et al., 2006; Kamangar et al., 2007). Esophageal cancer is a relatively common kind of cancer in the eastern countries (Delpisheh et al., 2014). Half million deaths is attributable to esophageal cancer and its five year survival rate is less than 10 percent (Ma and Yu, 2006; Fauci et al., 2008; Rasouli et al., 2011). Previous epidemiological studies in Asia have shown considerable changes in the incidence of esophageal cancer. Iran is located in areas of Asia with high rates of incidence (Long et al., 2010; Moore et al., 2010a; Moore et al., 2010b; Salim et al., 2010). This cancer is more common in some geographical areas, known as the belt of esophageal cancer, that covers areas from north of China, west and center of Asia to the north of Iran (Kamangar et al., 2007). This suggests the existence of some risk factors with a heterogeneous
geographical distribution that are related to the incidence pattern (Corley and Buffler, 2001; Stein et al., 2005).

Incidence of esophageal cancer in Iran is higher than the global rate in both sexes (Moore et al., 2010b; Radmard, 2010). with age standardized incidence rate of 0.88 and 6.15 for females and males having the fifth and eighth ranks among all cancer sites respectively (E’temad and Gooya, 2011).

Obtaining a geographical pattern for disease has been used since the early decades of the 20th century both for communicable and non-communicable diseases. Its use, nowadays, has been developed so that health designers apply it as a way to do more qualified interventions and help disease prevention. On the other hand, disease registration based on geographical regions, makes it possible to test variability patterns. It is also useful to recognize low and high risk regions (Lawson et al., 2001).

A number of studies have been conducted about the mapping of GI tract cancers in Iran (Mahaki et al., 2011; Asmarian et al., 2012; Asmarian et al., 2013a; Asmarian et al., 2013b; Mahaki et al., 2013). But few investigations have been done about the mapping of esophageal cancer in recent years (Poor Ahmad and Yavar, 2002; Asmarian et al., 2013c).

Incidence studies are descriptive and have been conducted for specific individual years or province. None of them were at the province and county levels and/or for a wider range of time.

Due to the high incidence of esophageal cancer and necessity of providing knowledge about its geographical distribution at province and county levels, present study aims to use Bayesian multilevel space-time models to obtain more accurate estimates and maps for relative risk of esophageal cancer.

The merits of this model include the power of tolerance for missing data without need of estimation and loss of data and the mean studied population, the possibility of considering the role of neighboring areas in different levels, the possibility of determining the relative effectiveness of every level hierarchically over the response variable and, finally, the feasibility of determining effective factor over the response variable for each levels. These models provide more precise estimates and more clear responses in comparison with single level analyses (Hosseini and Gohari, 2014).

Materials and Methods

This is an ecological study conducted in Iran in county level. The data were obtained from reports by Non-communicable Diseases Center of Health and Medical Education Ministry’s Diseases Management Center (http://www.irccancer.ir). On esophageal cancer in 320 counties (ICD-10) from 2005 to 2007. The census of the year 2006 and also estimated population of Iranian Statistics Center were used for the population which is potentially at risk (Iran, 2012).

The two geographically nested levels of counties and provinces were nested in time episodes. The expected number of esophageal cancer cases was calculated on the basis of population size and the observed number of cancer cases during this period. In order to prevent assessing geographical distribution from being distorted by other factors, we included risk factors, including physical activity, smoking, fruit and vegetable consumption and weight at province level and obtained adjusted estimates. Smoking was considered as multiplication of the population percent of smokers in the mean number of used cigarettes per day in every province. Also, obesity was regarded as the percent of the population having BMI>25 in every province. Fruit and vegetable consumption were measured as the total of mean fruit and vegetable number used daily in every province. Physical activity was calculated using a combined index called Metabolic Equivalent (MET).

For rare diseases such as cancers, it is usually assumed that the number of patients follows a Poisson distribution in every region. For modeling the data, multilevel model was used where parameters, in every level, can be random and the function of the characteristics of that level or other levels. One variation of the multilevel model is multiple-membership model as follows.

\[
\begin{align*}
\log(\lambda) &= \text{Log} (\sigma_i) + X \beta + \sum_{j=1}^{n_i} w_{ij} (\delta) + u_i^{(j)} \\
\end{align*}
\]

Where \([i]\) is the county indicator and random effects are defined by \(u\) and \(\delta\) represent corresponding level. Since number 1 is supposed for observation level, the numbering process is started from 2. There are regional and neighborhood effects indicated by number 2 and 3, respectively. Here \(\beta\) and \(x\) represent fixed effects and design matrix.

Regions \([i]\) refer to counties and neighbor \([j]\) denote the set of neighboring areas of each county. Therefore, in this model, the specification of the observed count is affected by different predictor variables. Typically the weights in this model are such that \(\sum_{j=\text{neighbor}(i)} w_{ij}(\delta)=1\) and, generally, all neighbors are given equal weights so that, in fact, \(w_{ij}(\delta) = \frac{1}{R}\) where \(n_i\) is the number of neighbors to Region \([i]\). Bayesian estimation is used for fitting the model (Lawson et al., 2003).

The full Bayesian method is one of the common ways in diseases mapping. Structural and non-structural heterogeneities are regarded by this method. We calculate relative risk of esophageal cancer by use of full Bayesian model suggested by Besag et al. known as BYM model from 2005 to 2007. This spatial model has no parameter to capture the effect of time (Lawson et al., 2003).

We also used Deviance Information Criterion (DIC) and Logarithmic Score (LS) to compare and find the best model. The model with the least LS and DIC is supposed to have the highest quality of prediction among all models. DIC is applied to assess the goodness of fit of Bayesian models and is calculated as \(DIC = D + pD = \hat{D} + 2pD\) where, \(D, \hat{D}\) and \(pD\) represent the mean deviance posterior distribution, estimation of pointed variation and the number of effective parameters of the model, respectively. \(pD\) is found by \(pD = \hat{D} - D\) As a result, a model with a best trade-off between DIC, LS, and \(pD\) would be considered as the best fit for the data (Spiegelhalter et al., 2002; Celeux et al., 2006).
In this study, analysis was performed using INLA Bayesian approach in R software. INLA approach is used as an effective calculating method. This method is not only faster than McMC methods, but also does not have the challenges related to convergence and other assumptions of McMC (Schrödle and Held, 2011; Blangiardo et al., 2013).

Results

There were the total numbers of 7160 esophageal cancer cases in Iran in 2005-2007. The highest number of cases was in the Mashhad with 684 cases in a three-year period. Table 1 shows the DIC, pD, and LS for different models. The description for each model is stated in the table.

Model 6 has the least amount of DIC, LS and fits the data more appropriately. The results of these models are shown in Table 2. In the model 6, the effect of year was significant and increasing. Physical activity was also significant. The same results were happened in model 7. But the other factors were not significant.

Figure 1, a, b and c shows the relative risk of the counties by adjusting risk factors, including smoking and physical activity. In model 6, the effect of year was significant and increasing. Physical activity was also significant. The same results were happened in model 7. But the other factors were not significant.

Table 1. Results Indicators of DIC, PD, D, LS Models are Fitted To Compare

<table>
<thead>
<tr>
<th>Model</th>
<th>DIC</th>
<th>PD</th>
<th>D</th>
<th>LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: without adjustment for risk factors</td>
<td>4345</td>
<td>310</td>
<td>4035</td>
<td>2.213</td>
</tr>
<tr>
<td>Model 2: adjusted risk factor for smoking</td>
<td>4154</td>
<td>216</td>
<td>3938</td>
<td>2.26</td>
</tr>
<tr>
<td>Model 3: adjusted risk factor for fruits and vegetables</td>
<td>4347</td>
<td>310</td>
<td>4036</td>
<td>2.316</td>
</tr>
<tr>
<td>Model 4: adjusted risk factor for physical activity</td>
<td>4114</td>
<td>216</td>
<td>3895</td>
<td>2.231</td>
</tr>
<tr>
<td>Model 5: The adjusted risk weight</td>
<td>4157</td>
<td>216</td>
<td>3941</td>
<td>2.263</td>
</tr>
<tr>
<td>Model 6: with adjustment for risk factors of smoking and physical activity</td>
<td>4111</td>
<td>218</td>
<td>3894</td>
<td>2.231</td>
</tr>
<tr>
<td>Model 7: by discounting all factors</td>
<td>4114</td>
<td>218</td>
<td>3895</td>
<td>2.236</td>
</tr>
</tbody>
</table>

Table 2. Estimated Coefficients of Risk Factors for Models Fitted

<table>
<thead>
<tr>
<th>Parameter estimation</th>
<th>95% CrI</th>
<th>Parameter estimation</th>
<th>95% CrI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.4071</td>
<td>(0.1352, 0.6800)</td>
<td>0.5333</td>
</tr>
<tr>
<td>Smoking</td>
<td>-0.0012</td>
<td>(-0.0025, 0.0001)</td>
<td>-0.0013</td>
</tr>
<tr>
<td>Weight</td>
<td>-0.0029</td>
<td>(-0.0147, 0.0088)</td>
<td>-0.0244</td>
</tr>
<tr>
<td>Physical activity</td>
<td>-0.0243</td>
<td>(-0.0315, -0.0171)</td>
<td>0.0056</td>
</tr>
<tr>
<td>Fruit and vegetable</td>
<td>0.0836</td>
<td>(0.0462, 0.1210)</td>
<td>0.0859</td>
</tr>
<tr>
<td>Year</td>
<td>0.0836</td>
<td>(0.0462, 0.1210)</td>
<td>0.0859</td>
</tr>
</tbody>
</table>

Figure 1. Relative Risk of Incidence Esophageal Cancer in Counties. a) multilevel model by adjusting risk factors including smoking and physical activity in 2005; b) multilevel model by adjusting risk factors including smoking and physical activity in 2006; c) multilevel model by adjusting risk factors including smoking and physical activity in 2007; f) multilevel model by adjusting all studied risk factor in 2005; g) multilevel model by adjusting all studied risk factor in 2006; d) multilevel model by adjusting all studied risk factor in 2007

DOI:http://dx.doi.org/10.7314/APJCP.2015.16.14.5787
Table 3. ICC values

<table>
<thead>
<tr>
<th>Level</th>
<th>Variance</th>
<th>ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Province</td>
<td>0.366</td>
<td>57%</td>
</tr>
<tr>
<td>County</td>
<td>0.267</td>
<td>42%</td>
</tr>
</tbody>
</table>

Discussion

According to the results, north-east and northwest regions and some parts of central zones (especially Khorasan, North Khorasan, Golestan, Ardabil, West Azerbaijan, East Azerbaijan, Kurdistan and some regions of Semnan provinces) have higher levels of esophageal cancer incidence (relative risk more than 3.5) in comparison with desert and southern regions (especially Sistan and Baluchestan, Bushehrand and Hormozgan provinces). It is necessary to investigate the causes of cancer in high-risk areas and the relationship between these factors and the results of this study.

The present investigation is the first mapping of esophageal cancer incidence rate in Iran at county level. Esophageal cancer incidence and related mortality in Iran is one of the important health problems and climate conditions facilitates its incidence and prevalence of esophageal cancer. Hence in the present study, environmental investigation and its mapping were done and relative risks were estimated in the counties during three years. Esfandiari Islami et al., 2009a; Islami et al., 2009b) indicated the importance of the Sabalan volcano as an important environmental factor in the incidence of esophageal cancer in Ardebil province (Ahari et al., 2013).

As in the work by Azami et al. (2006) found no statistically significant association between smoking and vegetable intake and esophageal cancer incidence. In the investigation of Hajizadeh et al. (2012) there was no correlation between vegetable intake and esophageal cancer as well. In contrast to the results of previous studies (Lagergren et al., 1999; Lagergren et al., 2014; Thrift et al., 2014), weight was not significant factor in esophageal cancer incidence in our study. Physical activity has significant effect, which is in accordance with the results of previous studies (Chen et al., 2014; Dos Santos et al., 2014; Thrift et al., 2014), weight was not significant factor in esophageal cancer incidence in our study. Physical activity has significant effect, which is in accordance with the results of previous studies (Chen et al., 2014; Dos Santos et al., 2014; Thrift et al., 2014). This risk factor may be accompanied by decreasing incidence risk of esophageal cancer.

Many articles have been published in Iran about esophageal cancer in which this kind of cancer has been studied in a single or a few provinces especially in Northern provinces. They are, mostly, descriptive and cross-sectional (Ghavamzadeh et al., 2001; Radmard, 2010; Sadjadi et al., 2010; Rasouli et al., 2011; Rahimzadeh Bozorgi et al., 2013; Somi et al., 2014). We found no study in which environmental factors have been taken into consideration the esophageal cancer incidence nationwide. The specialists and health planners should present their beliefs for designing the effective programs, in order to study precisely influencing factors in the incidence of esophageal cancer and the sort of achieved pattern in the variation of this cancer, especially the approach of facing high risk areas.

It should be noted that just pathological reports are
taken into account in an Iranian registration system of cancer. So, at best, only 80 percent of cancer cases can be registered. In addition, this registration system was disqualified in some counties and years so that it apparently leads to underestimate the number of cancer cases (Asmarian et al., 2013a). Mapping of a disease at county level will be smoother than province level due to more areas in county level mapping.

Because access to up-to-date data about esophageal cancer is not available, other types of risk factors such as tobacco consumption, alcohol use, warm drinks like tea, socioeconomic condition, unhealthy nutrition, Selenium, shortage, viral and genetic agents were not registered, and also the data related to effective risk factors in county level were not registered and there are missed data in county level. Using more precise and complete data will help further clarifying the process of esophageal cancer in Iranian counties. It is also recommended to increase the period of study or to investigate the impact of each factor in every region by use of varying slope models. To find more clearly risk factors for esophageal cancer incidence, it is suggested conducting further studies in the areas with high incidence.

References


