CrossMark

ELSEVIER

ORIGINAL ARTICLE

Gastric and Esophageal Cancers Incidence Mapping in Golestan Province, Iran: Using Bayesian—Gibbs Sampling

Atefeh-Sadat Hosseintabar Marzoni [a], Abbas Moghimbeigi [b,*], Javad Faradmal [b]

[a]Golestan Research Center of Gastroenterology and Hepatology, Golestan University of Medical Sciences, Golestan, Iran.
[b]Modeling of Noncommunicable Diseases Research Center, Department of Biostatistics and Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran.

Abstract

Objectives: Recent studies of esophageal cancer (EC) and gastric cancer (GC) have been reported to have high incidence rates of these cancers in Golestan Province of Iran. The present study describes the geographical patterns of EC and GC incidence based on cancer registry data and display statistically significant regions within this province.

Methods: In order to map the distribution of upper gastrointestinal cancer, relative risk (RR) were calculated. Therefore, to estimate a more reliable RR, Poisson regression models were used. The adjusted models (adjusted to urban—rural area, sex, and grouped age proportion) were utilized. We considered two-component random effects for each observation, an unstructured (non-correlated) and a group of "neighbor" (correlated) heterogeneities. We estimated the model parameters using Gibbs sampling and empirical Bayes method. We used EC and GC data that were registered with Golestan Research Center of Gastroenterology and Hepatology in the years 2004—2008.

Results: The EC and GC maps were drawn for 2004—2008 in the province. Kalaleh and Minoodasht counties have a high RR of EC and GC in the years of study. In almost all years, the areas with a high RR were steady.

Conclusion: The EC and GC maps showed significant spatial patterns of risk in Golestan province of Iran. Further study is needed to multivariate clustering and mapping of cancers RRs with considering diet and socioeconomic factors.

*Corresponding author.
E-mail: moghimb@yahoo.com, moghimbeigi@umsha.ac.ir (A. Moghimbeigi).

1. Introduction

About 70,000 new cases of cancer were diagnosed in the Iranian population in 2008 [1]. Esophageal cancer (EC) is the second and the third most common malignancy in Iranian men and women, respectively [2]. Also, gastric cancer (GC) is a major problem in the world and it is the second leading cause of cancer deaths [3]. The incidence rate of EC/GC is 6.25/8.89 and 5.83/15.21 for women and men, respectively (during the period 2005−2006) [4]. The northeastern part of Iran is known as the high-risk regions of the EC and GC in both male and female sexes [5]. Golestan is one of the northern provinces of Iran. The age-standardized incidence rate (ASR) per 100,000 person−years of EC in Gonbad (a county in Golestan, Iran) is > 100 and this city is one of the high risk areas in the world [2]. The estimations of ASR (per 100,000) of EC are 17.6 and 14.4 in Iran [6] and 43.3 and 36.3 in the province for men and women, respectively [2]. Preliminary research carried out by the Iran cancer institute has shown that EC accounts for about 9% of all cancers and 27% of digestive cancers, and its prevalence in men is about 1.7 times higher than in women [4]. Recent research has reported that ASR (per 100,000) of GC in Iran is about 26.1 in and 11.1 in women [6].

In recent years, several studies have been conducted to map the geographical spread of EC and GC incidence using adjusted age-specific standardized incidence ratio (SIR) in the southwest of the Caspian Sea from 2001 to 2005 [7,8]. The previous studies have considered counties of northern provinces of Iran as clustered. Most counties of Golestan are in the high-risk incidence rate cluster [7,8]. However, the incidences of these cancers are not the same in all regions of the province, despite being in a high-risk cluster. The purpose of this study is to evaluate and adjust EC and GC for contextual risk factors from 2004 to 2008, then identify counties in Golestan province that have the highest observed count/expected count of these cancers compared with other regions within this province. Comparison with the results of other studies over the years can also be valuable.

2. Materials and methods

2.1. Study population

The population this study was residents of Golestan province. The estimated midyear population between 2004 and 2008 that are stratified by sex, age ($\leq$ 69 years and > 69 years), and place of residence (urban or rural) was obtained from the statistical center of Iran. The occurrences of new cases of EC and GC during a period of 4 years (2004−2008) were established from Golestan Research Center of Gastroenterology and Hepatology. The cancers were registered with procedures that are widely established throughout the world by the International Agency for Research on Cancer, the International Association of Cancer Registries, and the World Health Organization.

2.2. Statistical analysis

In many studies, the response variable is the counts of rare events, such as the number of new cancer cases in the population during a specified time. In such cases it is assumed that the response variable has a Poisson distribution [9]. In this study, SIR, the ratio of observed new cases (y_i) to the expected number of new cases (e_i) was used as the response variable.

$$\widehat{\theta}_i = SIR_i = \frac{y_i}{e_i}$$

The SIR is a crude estimate of underlying regional-specific relative risks (RR). Hence smoothed estimates of RR for disease mapping were calculated using empirical Bayes method. We suppose that y_i, the number of disease observed in study i-th county, has a Poisson distribution [$y_i \sim$ Poisson $(\theta_i e_i)$, i = 1,…,N]. where θ_i is the RR and the expected number of cases in the i-th county calculates as:

$$e_i = n_i \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} n_i}$$

where n_i is the population at risk in the i-th county.

In the model for RR to account heterogeneity, we considered two-component random effects for each observation, an unstructured (noncorrelated heterogeneity) and a group of "neighbor" (correlated heterogeneity) random effects [10−12]. This model has been presented and extended for disease mapping and clustering [13,14]. It is formulated as follows:

$$\log \theta_i = \alpha + u_i + v_i$$

where α is the overall effect, u_i is correlated heterogeneity, and v_i is the uncorrelated heterogeneity. Whereas estimating RR in each region depends on the neighborhood, we applied the clustering structure for the spatial correlations. Where $v_i \sim N(0, \tau_v^2)$ and accordingly for u_i, Besag and Newell [14] have proposed a conditional autoregressive structure as:

$$\left[u_i | u_j, i \neq j, \tau_u^2\right] \sim N\left(\overline{u}_i, \tau_i^2\right)$$

$$\overline{u}_i = \frac{1}{\sum_j w_{ij}} \sum_j u_j w_{ij}$$

$$\tau_i^2 = \frac{\tau_u^2}{\sum_j w_{ij}}$$

$$w_{ij} = \begin{cases} 1 & \text{if } i,j \text{ counties are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

Table 1. The median (M), and 2.5 and 97.5 percentiles of covariates effects on log (relative risk) in the adjusted model.

Variables		2004 2.5%	M	97.5%	2005 2.5%	M	97.5%	2006 2.5%	M	97.5%	2007 2.5%	M	97.5%	2008 2.5%	M	97.5%
Area	EC	−0.960	−0.007	1.196	−0.614	0.284	2.327	−0.881	0.050	1.523	−0.995	−0.043	0.931	−0.709	0.139	1.085
	GC	−0.893	0.006	0.914	−0.697	0.058	0.934	−0.907	−0.034	0.817	−0.795	0.030	1.066	−1.037	−0.079	0.747
Sex	EC	−2.328	−0.038	2.183	−2.888	0.042	3.368	−2.302	−0.019	2.535	−2.323	−0.040	2.139	−2.391	0.004	2.201
	GC	−2.203	−0.017	2.050	−2.092	0.001	2.124	−2.093	−0.014	2.091	−2.074	−0.011	2.197	−2.238	−0.033	2.032
Age	EC	−1.980	−0.067	1.825	−2.702	−0.108	2.588	−2.117	−0.068	1.979	−1.704	−0.022	1.974	−2.228	−0.062	2.172
	GC	−1.749	−0.010	1.718	−1.736	−0.020	1.706	−1.626	−0.003	1.747	−1.973	−0.052	1.635	−1.735	−0.015	1.732

EC = esophageal cancer; GC = gastric cancer.

where, the τ_u and τ_v control variability of u and v. We have used the empirical Bayes model for estimating the model parameters. Uninformative prior distributions for the parameters of the model were considered as:

$$\alpha \sim N(0, 0.001)$$

$$\tau_u \sim gamma(0.01, 0.01)$$

$$\tau_v \sim gamma(0.01, 0.01)$$

The RR adjusted for the effect of the residential area (rural = 0; urban = 1), sex (male = 0; female = 1), and grouped age (< 69 years = 0;   69 years = 1) as:

$$\log \theta_i = \beta_1 + \beta_2 area[i] + \beta_3 gender[i] + \beta_4 age[i] + u_i + v_i$$

We used OpenBUGS version 3.1.2 (produced by Medical Research Council (MRC) and Imperial College, UK), the Bayesian analysis of complex statistical software to estimate parameters of the model with the Gibbs−Bayesian method. We considered two independent Markov chains. To ensure the convergence of chains, after visual inspections, we used Gelman−Rubin, Geweke, Raftery−Lewis, and Heidelberg−Welch diagnostic tests via R using the coda package [12,15]. After a sufficient (10,000) burn-in to remove the effects of the initials, the following 100,000 iterations were sampled from each of the two chains choosing lag = 10 to avoid possible autocorrelation. We considered multivariate normal as a prior for β_1, β_2, β_3, and β_4 parameters. The estimated RRs were subsequently mapped with the GeoBugs tool in the OpenBUGS version 3.1.2.

Table 2. The median (M), and 2.5 percentile and 97.5 percentile of relative risk in each ward.

County		2004 2.5%	M	97.5%	2005 2.5%	M	97.5%	2006 2.5%	M	97.5%	2007 2.5%	M	97.5%	2008 2.5%	M	97.5%
Kalaleh	EC	1.638	2.211	2.893	1.269	1.763	2.359	1.470	2.010	2.690	1.223	1.703	2.317	1.321	1.806	2.425
	GC	0.883	1.247	1.751	0.830	1.143	1.582	0.855	1.202	1.669	0.811	1.169	1.649	0.805	1.152	1.618
Gonbad	EC	0.917	1.216	1.577	1.000	1.313	1.704	0.974	1.290	1.674	0.852	1.122	1.443	0.990	1.282	1.501
	GC	0.951	1.235	1.604	0.758	0.983	1.254	0.788	1.029	1.326	0.635	0.872	1.147	0.659	0.895	1.174
Minoodasht	EC	0.933	1.354	1.915	0.934	1.362	1.930	1.163	1.661	2.309	0.775	1.148	1.641	1.037	1.485	1.795
	GC	0.661	0.994	1.412	0.670	0.956	1.315	0.880	1.234	1.762	0.838	1.215	1.752	0.984	1.415	2.028
Aghghala	EC	0.303	0.578	0.961	0.719	1.120	1.686	0.447	0.782	1.253	0.460	0.762	1.158	0.462	0.776	1.017
	GC	0.409	0.733	1.115	0.713	1.014	1.423	0.532	0.845	1.211	0.471	0.793	1.206	0.518	0.829	1.228
BandarTurkaman	EC	0.384	0.656	1.039	0.293	0.547	0.881	0.565	0.909	1.412	0.409	0.685	1.041	0.413	0.680	0.921
	GC	0.455	0.751	1.101	0.587	0.878	1.206	0.507	0.795	1.133	0.385	0.682	1.052	0.490	0.787	1.148
Azadshahr	EC	0.523	0.901	1.442	0.517	0.891	1.425	0.407	0.770	1.277	0.585	0.949	1.448	0.475	0.832	12.212
	GC	0.680	1.045	1.556	0.535	0.857	1.213	0.670	1.015	1.492	0.466	0.807	1.256	0.618	0.977	1.467
Ramian	EC	0.213	0.505	0.933	0.331	0.694	1.183	0.209	0.511	0.958	0.376	0.709	1.132	0.341	0.687	0.929
	GC	0.521	0.888	1.333	0.480	0.809	1.166	0.378	0.748	1.129	0.565	0.924	1.447	0.446	0.789	1.219
Aliabad	EC	0.416	0.694	1.073	0.399	0.700	1.074	0.496	0.808	1.252	0.406	0.688	1.034	0.596	0.911	1.220
	GC	0.780	1.114	1.605	0.636	0.921	1.245	0.739	1.046	1.482	0.353	0.637	0.997	0.465	0.765	1.118
Gorgan	EC	0.507	0.701	0.941	0.479	0.665	0.898	0.365	0.531	0.739	0.753	0.974	1.250	0.462	0.635	0.811
	GC	0.606	0.812	1.054	0.842	1.065	1.331	0.762	0.972	1.233	0.945	1.202	1.505	0.882	1.124	1.411
Kordkooy	EC	0.449	0.821	1.403	0.501	0.892	1.522	0.280	0.599	1.096	0.324	0.643	1.086	0.415	0.763	1.114
	GC	0.555	0.919	1.424	0.732	1.066	1.592	0.506	0.847	1.285	0.828	1.311	2.090	0.395	0.745	1.181
Bandar Gaz	EC	0.343	0.717	1.372	0.208	0.540	1.042	0.233	0.575	1.145	0.317	0.668	1.194	0.471	0.886	1.232
	GC	0.674	1.114	1.916	0.618	0.984	1.524	0.406	0.781	1.258	0.437	0.851	1.454	0.475	0.872	1.473

EC = esophageal cancer; GC = gastric cancer.

Figure 1. Geographic boundaries of wards in Golestan provinces.

3. Results

The total numbers of new cases of the EC and GC during the study period (2004−2008) were 1100 and 1122, respectively. The mean (standard deviation) age of EC and GC cases were 64.96 (11.919) and 65.16 (12.975), respectively. 56.6% of EC and 70.9% of GC cases were male and 32.7% of EC and 42.2% of GC cases were resident in urban areas.

Adjustment models by area (urban−rural), sex, and age effects to log (RR) fitted. Table 1 shows 2.5%, median and 97.5% of posterior distributions of model parameters in the years. This reveals that none of the covariates significantly affects the (logarithm of) RR of EC and GC.
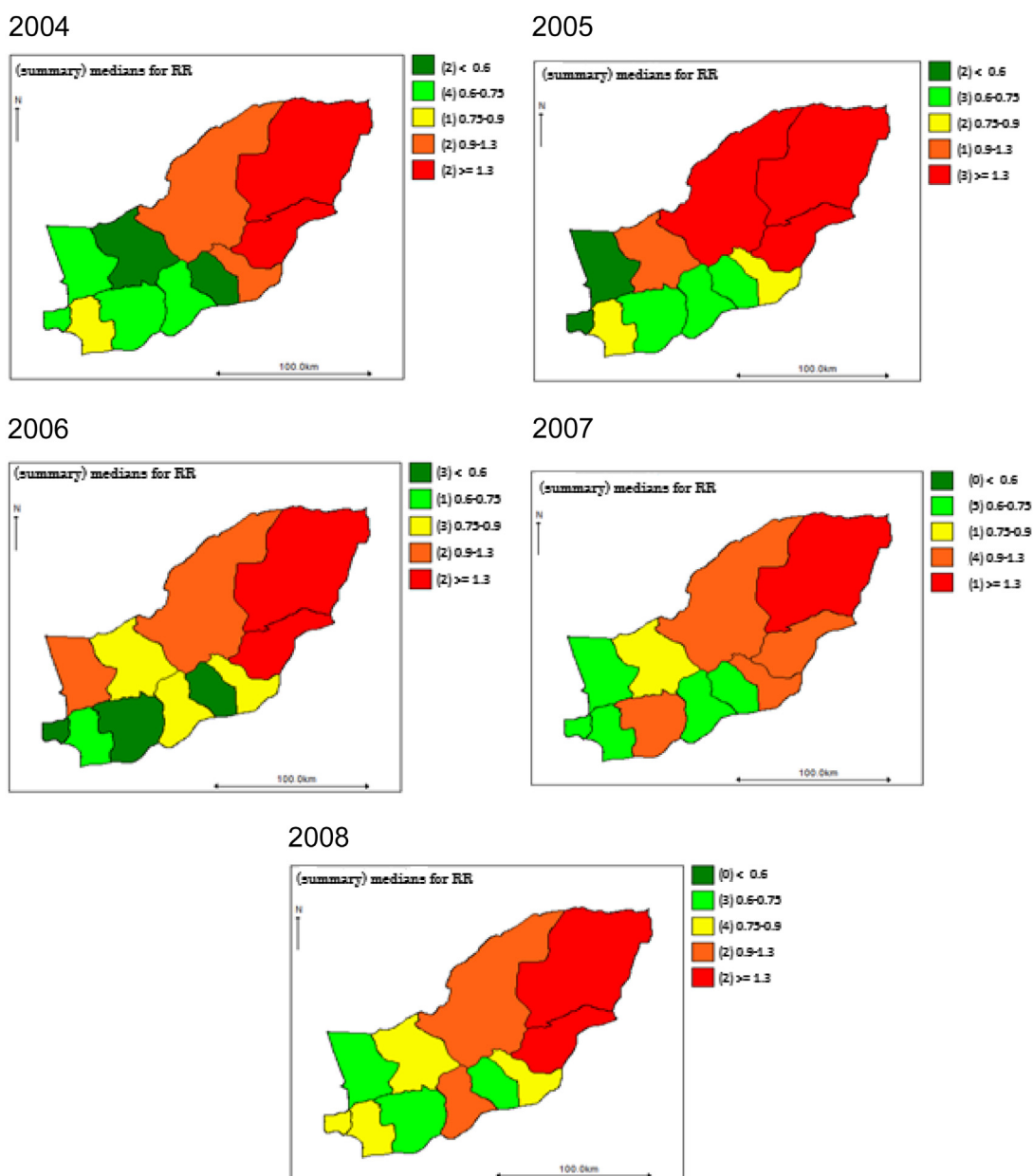

Figure 2. Spatial pattern, local clusters and smoothed relative risk of esophageal cancer incidence during 2004−2008.

Table 2 shows the median, and 2.5 percentile and 97.5 percentile in the RR (observed counts/expected counts) of diseases. In Kalaleh, the RR of EC is higher than other counties, such that the 2.5 percentile of the RR is > 1 in every year. Figure 1 shows geographic boundaries of the province. As adjusted models show (Table 2) the covariates of study are not significant and when we mapped RR of GC and EC separately for both genders, the results were similar. We decided to map RR separately in years. Figures 2 and 3 show geographical patterns of EC and GC in the province. In these figures

the counties were clustered into four groups; counties with: RR < 0.6, 0.6   RR < 0.75, 0.75   RR < 0.90, 0.9   RR < 1.3, and RR > 1.3. It is clearly shown that Kalaleh, Minoodasht, and Gonbad have a higher RR of EC and GC than other counties in almost all years.

4. Discussion

In this study, we have created separate maps of EC and GC disease in the counties of Golestan province during 2004−2008. For comparison of incidence rates in
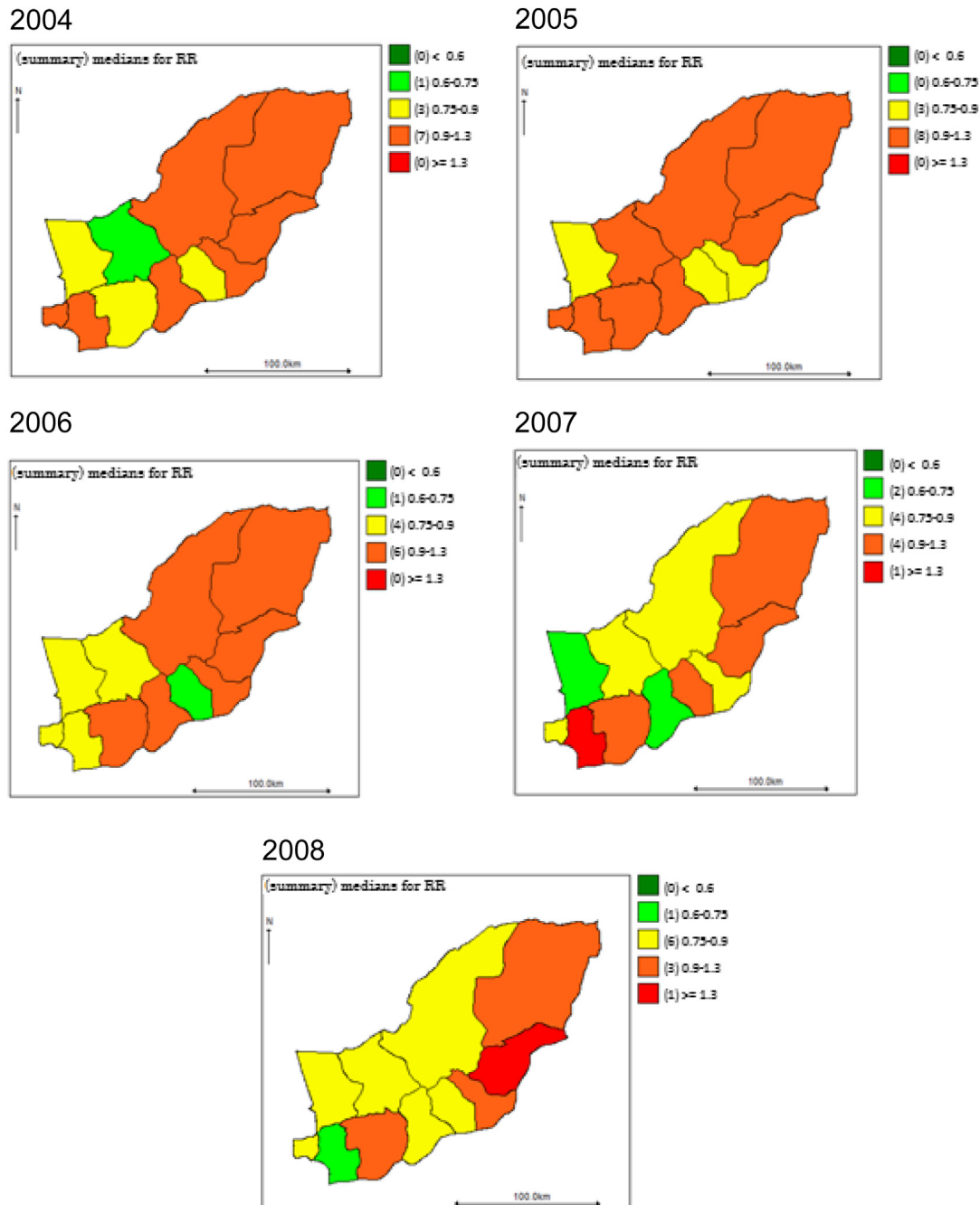

Figure 3.   Spatial pattern, local clusters and smoothed relative risk of gastric cancer incidence during 2004−2008.

the counties, we adjusted RR with sex (male, female), age ($<$ 69 years, $\geq$ 69 years) resident in the region (rural, urban) populations. Based on the results of modeling and mapping for RR of EC and GC, it has been observed, the RR is higher for some counties and it highlights the possible role of geographical and cultural differences. According to Figure 2, Kalaleh and Minoodasht (except in 2007) have experienced a higher risk of incidence of EC than expected. After these two counties, Gonbad has a high RR of EC. In the previous study in the Caspian region of Iran (Mazandaran and Golestan provinces) using data collected during 2001−2005, Kalaleh, Minoodasht, Azadshahr, and Gonbad have higher risk of incidence of EC ($\geq$ 1.33 times the expected) in both sexes [7]. In a recent study, adjusted on a diet and socioeconomic factors, it was found that all regions of Golestan province have high standardized incidence rates of EC/GC [8]. We found that Gonbad has the highest; and Kalaleh and Minoodasht have a high RR of GC in both genders. Also, almost in all years, the areas with a high RR have been steady.

Our study investigated the geographical distributions of new GC/EC cases to determine the counties at a high risk of cancer incidence. Our results are somewhat similar to the previous studies that include this province or neighborhood provinces [7]. Mohebbi et al [8] have adjusted the RR of EC and GC to diet and socioeconomic factors in the province and neighboring areas. Multivariate clustering and mapping of cancers RRs with considering diet and socioeconomic factors are interesting for future studies.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

1. Ramezani Gourabi B. Recognition of geographical diffusion esophagus cancer in southwestern of Caspian Sea. Am J Sci 2011 Feb 25;7(2):297−302.
2. Sadjadi A, Marjani H, Semnani S, et al. Esophageal cancer in Iran: a review. Mid E J Cancer 2010 Jan;1(1):5−14.
3. Kamangar F, Dores GM, Anderson WF. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. J Clin Oncol 2006 May 10;24(14):2137−50.
4. Kolahdoozan S, Sadjadi A, Radmard A, et al. Five common cancers in Iran. Arch Iran Med 2010 Mar;13(2):143−6.
5. Mahboubi E, Kmet J, Cook PJ, et al. Oesophageal cancer studies In the Caspian Littoral of Iran: the Caspian cancer registry. Br J Cancer 1973 Sep;28(3):197−214.
6. Sadjadi A, Nouraie M, Mohagheghi MA, et al. Cancer occurrence in Iran in 2002, an International perspective. Asian Pacific J Cancer Prev 2005 Jul−Sep;6(3):359−63.
7. Mohebbi M, Mahmoodi M, Wolfe R, et al. Geographical spread of gastrointestinal tract cancer incidence in the Caspian Sea region of Iran: Spatial analysis of cancer registry data. BMC Cancer 2008 May 14;8:137.
8. Mohebbi M, Wolfe R, Jolley D, et al. The spatial distribution of esophageal and gastric cancer in Caspian region of Iran: An ecological analysis of diet and socio-economic influences. Int J Health Geogr 2011 Feb 15;10:13.
9. Agresti A. An introduction to categorical data analysis. Second ed. John Wiley& Sons; 2007.
10. Lawson A, Browne W, Rodeiro C. Disease Mapping with Winbugs and MLwiN. John Wiley & Sons; 2003.
11. Lawson A, Williams F. An introduction guide to disease mapping. John Wiley& Sons; 2001.
12. Ntzoufras I. Bayesian modeling using WinBUGS. Wiley; 2009.
13. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 1987 Sep;43(3):671−87.
14. Besag J, Newell J. The detection of clusters in rare diseases. J R Stat Soc Series A Stat Soc 1991;154:143−55.
15. Plummer M, Best N, Cowles K, et al. CODA: Convergence diagnosis and output analysis for MCMC. R News 2006;6(1):7−11.