CrossMark

ORIGINAL ARTICLE

ELSEVIER

Joint Disease Mapping of Two Digestive Cancers in Golestan Province, Iran Using a Shared Component Model

Parisa Chamanpara [a,b], Abbas Moghimbeigi [b,*], Javad Faradmal [b], Jalal Poorolajal [b]

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

Abstract

Objectives: Recent studies have suggested the occurrence patterns and related diet factor of esophagus cancer (EC) and gastric cancer (GC). Incidence of these cancers was mapped either in general and stratified by sex. The aim of this study was to model the geographical variation in incidence of these two related cancers jointly to explore the relative importance of an intended risk factor, diet low in fruit and vegetable intake, in Golestan, Iran.

Methods: Data on the incidence of EC and GC between 2004 and 2008 were extracted from Golestan Research Center of Gastroenterology and Hepatology, Hamadan, Iran. These data were registered as new observations in 11 counties of the province yearly. The Bayesian shared component model was used to analyze the spatial variation of incidence rates jointly and in this study we analyzed the data using this model. Joint modeling improved the precision of estimations of underlying diseases pattern, and thus strengthened the relevant results.

Results: From 2004 to 2008, the joint incidence rates of the two cancers studied were relatively high (0.8—1.2) in the Golestan area. The general map showed that the northern part of the province was at higher risk than the other parts. Thus the component representing diet low in fruit and vegetable intake had larger effect of EC and GC incidence rates in this part. This incidence risk pattern was retained for female but for male was a little different.

Conclusion: Using a shared component model for joint modeling of incidence rates leads to more precise estimates, so the common risk factor, a diet low in fruit and vegetables, is important in this area and needs more attention in the allocation and delivery of public health policies.

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

1. Introduction

Cancer is the third leading cause of death and nearly 70,000 new cases of cancer occur annually in Iran [1,2]. About half of all cancers are related to the gastrointestinal cancers. In men, the three important cancers are gastric, esophagus, and colorectal; in women, after breast cancer, these three are the major cancers [3]. There is an evidence of sharp gradients in incidence rates of esophagus cancer (EC) and gastric cancer (GC) over proportionally short geographical distances in the Caspian region of Iran [4]. In this area, EC is the second highest cause of death after heart disease [2]. Also, among other tumors, GC had a strikingly similar incidence [5]. Some studies have highlighted a positive correlation between standardized incidence ratios of GC and EC which might be an evidence of these two cancer sites shared common risk factors such as diet low in fruit and vegetable intake, low socio-economic status, smoking, and gastric atrophy but in the Caspian sea region of Iran, the first two component were more influential [3].

In northeastern Iran, Golestan province is one of the very high-risk areas of EC in the world so that the rates are as high in women as in men in areas surrounding Gonbad, one of the major counties of Golestan province, Iran, and further to the East [6]. Recently in Iran, the age standardized incidence rate of EC and GC for men was about 17.6 per 100,000 person years and 26.1 per 100,000 person years and for women, were 14.4 and 11.1 [7,8].

In epidemiology, disease mapping has long been used in the statistical analysis of geographical variation of disease rate [9], which provides useful information such as describing areas of unusually high risk and assessment hypotheses, and producing a clean map of disease risk to allocate better resources and public health policies [10]. Mapping the population-based standardized mortality ratio or standardized incidence ratio, defined as the ratio of observed to expected count in the region under study, specified the situation of geographic dispersion of disease incidence and mortality rates [11]. Although these methods obtain unbiased estimators of relative risk (RR) but suffer from many problems: their variance is large in areas with a small population and small in areas with a large population; they do not differentiate between regions when there is no death; and they do not try to manifest any underlying structure in the data and are not parsimonious [10].

To remove these problems a variety of alternative models have been proposed. Among them, the Bayesian approach is suggested more because of the great flexibility in modeling options and a reliable output for inferential purposes. This approach considers spatial correlation of disease rates among neighboring areas to capture the geographical structure, so the estimates of the parameters in the model are more realistic [11].

Most of the studies in geographical modeling of diseases are based on a single disease, but because many diseases have common risk factors, recently joint disease mapping has appeared [12]. The definition of joint disease mapping is the spatial modeling of two or more diseases or the same disease in two or more subsets of the population at risk [11,13]. Joint modeling of different diseases has some advantages including the ability to assess shared and specific geographic patterns of risk among different diseases and improvement in the precision of estimation of underlying diseases pattern. Moreover, when interest is in a relatively rare disease, this model incorporates data from a more common, and related disease so strengthens the relevant results of the rare disease [13].

In recent decades, different methods have been proposed for joint disease mapping [14]. The first study that introduced joint spatial model analysis was done by Langford et al [15] and Leyland et al [16] whom used a multilevel model. Knorr-Held and Best [17] proposed a shared component model, then Held [18] extended a shared component model to analyze the spatial variation of several disease that allows the linear predictor to be decomposed into shared and disease-specific spatial variability components. In another study, joint modeling of two diseases applied using a proportional mortality model [13]. Moreover, in Manda et al's [19] study four joint modeling techniques were compared, including multivariate intrinsic conditional autoregressive model, multivariate multiple membership multiple classification model, shared-component, and proportional mortality models using EC and GC data. This article confirmed that the shared component model adds more versatility in answering more substantive epidemiological questions than the other three models [19].

Mohebbi et al [3,4] executed two studies in Caspian region of Iran included Golestan and Mazandaran provinces and presented the geographical patterns of EC and GC separately in this area. In both of them, Golestan was in high risk, especially for EC [3,4]. Therefore, the main object of the present paper is to apply a shared component model for joint modeling of EC and GC in Golestan province of Iran, for which diet low in fruit and vegetable intake is considered as a major risk factor, to explore the geographical variation of these two disease incidence rates. Also, we explore the differences of incidence rates between males and females by joint modeling of EC and GC separated by sex.

2. Materials and methods

Data on incident cases of EC and GC from 2004 and 2008 were extracted from Golestan Research Center of Gastroenterology and Hepatology. The cancers were

registered with procedures widely established throughout the world by the International Agency for Research on Cancer, the International Association of Cancer Registries, and the World Health Organization.

We calculated relative risk for each cancer site (with the number of expected cases calculated using the average number of cases per ward observed in Golestan province and the population in the 2006 census).

In this article, we applied the shared-component model to model the spatial variation incidence rates of the two cancers in which they share diet low in fruit and vegetables as a latent spatial component. We formulated the joint modeling described by Knorr-Held and Best [17] for the two-disease setting. The common feature of the shared-common model that we used is the latent component that act as surrogate for geographical variation of the unobserved spatially structured risk factor that affect two diseases.

Suppose that O_{ij} indicates that observed count for disease j in area i ($1 \leq i \leq 11$, j = 1,2) and E_{ij} presents the expected number of cases (as obtained by multiplying the overall incidence rate and the estimate of the ward population). O_{ij} follows Poisson distribution with mean $\mu_{ij} = E_{ij}.R_{ij}$ in which R_{ij} is the unknown parameter in the model. The maximum likelihood estimate of the incidence rate is obtained by dividing the observed count to expected count for cancer j in area i. As said before this estimation has some drawbacks, so to eliminate these problems we use the Besag-York-Mollié (BYM) model [20], which yields more reliable estimates for relative risk by borrowing information from neighboring areas.

In this model, the log of disease-specific area-level relative risks are decomposed into the sum of two components: unstructured and structured random effects. Unstructured random effect (uncorrelated heterogeneity) is a component that models the effects that vary between areas and we assume that it follows a normal distribution [$v_i \sim N(0, \tau_v^2)$]. Structured random effect (correlated heterogeneity) is a component that assumes local dependence in space; in other words it considers weight for neighboring areas. This component is modeled by the conditional autoregressive normal (CAR Normal) prior, which assumes that the conditional distribution of each area-specific spatially structured component, given all other spatial effects, is a normal distribution with mean equal to the average of its neighbors, and variance inversely proportional to the number of these neighbors, the more neighbors an area has, the greater the precision is for that area effect.

In this study we used Bayesian shared component model to analyze the spatial distribution of incidence rates of the two cancers jointly. We considered diet low in fruit and vegetable intake as a risk factor. Thus, we modeled the log relative risk as below:

$$\log(R_{i1}) = \alpha_1 + \lambda_i \delta_1 + \varepsilon_{i1}$$

$$\log(R_{i2}) = \alpha_2 + \lambda_i \delta_2 + \varepsilon_{i2}$$

Where R_{i1} is the log relative risk for EG and R_{i2} is the log relative risk for GC in ward i. The parameter α_j is the disease specific intercept and λ_i is the shared diet low in fruit and vegetable intake component common to both cancers in ward i. The contribution of the shared component to the overall relative risk is weighted by the scaling parameter δ to allow a different risk gradient (on the log-scale) to be the included terms. ε_{ij} are the disease specific heterogeneous effects to capture possible variations not explained by the terms included in the model [21].

For a Bayesian model, all unknown parameters, whether fixed or random effects, are given prior distributions. We need priors that combine the BYM framework to link risk in space. For the shared spatial random effects, λ_i, we assumed an intrinsic normal conditional autoregressive as a prior distribution with sum-to-zero constraints on the random effect terms. This was a spatially correlated distribution with unit weight for neighboring areas to capture local dependence in space. Moreover a flat prior was assigned to the cancer specific intercepts, αj. Independent normal prior distributions were used for the logarithms of the scaling parameters, log δ. We independently assigned a conjugate hyper-prior gamma (0.5, 0.0005) distribution [22] to the precision of the shared component, τ, which is weakly informative. Finally the disease-specific heterogeneity random effects, εij, were assigned a multivariate normal prior distribution with covariance matrix Σ to allow for correlations amongst the cancers. The inverse of this matrix known as a precision matrix, Σ-1 modeled to arise from a Wishart (Q,6) prior distribution, where Q is set to be a diagonal matrix with 1s [19,21].

The shared component model was fitted to data using full Bayesian estimation within WinBUGS version 3.2.2 software (MRC Biostatistics Unit, Cambridge, the United Kingdom). For the model, we used the first 30,000 draws as the burn-in period and then drew 15,000 more samples. After thinning by 15,we were left with 1000 samples to base posterior summaries upon. The iterations were sampled from each of the chains choosing every 10th iteration to avoid possible autocorrelation; we monitored all fixed effects, weight and variance parameters for convergence. We used the CODA R package for convergence diagnostic and output analysis. As a result, the Brooks−Gelman−Rubin and Geweke diagnostic tools confirmed rapid convergence by 45,000 and we based inference on 45,000 iterations for each of the two chains for posterior summaries [23].

3. Results

Based on the 2006 census in Iran, the total population of Golestan province was 1,617,087 persons. The minimum number of people in a county (Bandar Gaz) was

46,226 and the maximum (Gorgan) was 401,399. According to Golestan Research Center of Gastroenterology and Hepatology, 1100 cases of EC and 1087 cases of GC have been recorded from 2004 to 2008.

Our analysis is related to the incidence rates of EC and GC from 2004 to 2008. The result reported the relative risk estimates of these two cancers jointly with diet low in fruit and vegetables as a shared component. Moreover, we present the joint modeling of EC and GC in men and women, separately. Figure 1A displays the overall posterior median relative risk surface of joint analysis for EC and GC from 2004 to 2008. It can be seen that this map is composed of two colors, pink and yellow, which means the incidence rate is 0.8−1.2. Based on this plot, we can say the incidence rate of the northern half of the area is more than one. This part included the counties Kolaleh, Gonbad Kavoos, Minoodasht, Azadshahr, and Ramiyan. Figure 2C represents the posterior median relative risk surface of joint analysis for women, which has the same pattern as the general map. However, for men the distribution of incidence rate is a little different as shown in Figure 2B. This figure shows that the incidence rate of EC and GC appear to be relatively distributed across the region, found in the northeast, southeast, and southwest parts of the province. These parts included Kolaleh, Azadshahr, Ramian, Kordkuy, and Bandar Gaz counties. In summary, the dominant feature of the general joint map is an increasing trend from the southwest to the northeast.

4. Discussion

In this paper, the main object was using the share component model to analyze the joint spatial distributions of EC and GC incidence rates from 2004 to 2008. We specified the advantages of spatial analysis of disease rates, the purpose of joint modeling of different diseases and its benefits, the shared component model structure, assumptions and formulation, and the data sources.

In the model under consideration, we have included two cancer rates as response variables in relation to a diet low in fruit and vegetables, as a risk factor, which is shared by cancers.

The resulting maps showed the geographical differences in cancers incidence rates and high risk areas in the target province. As we have seen, the general joint map showed that the northern half of the province was at
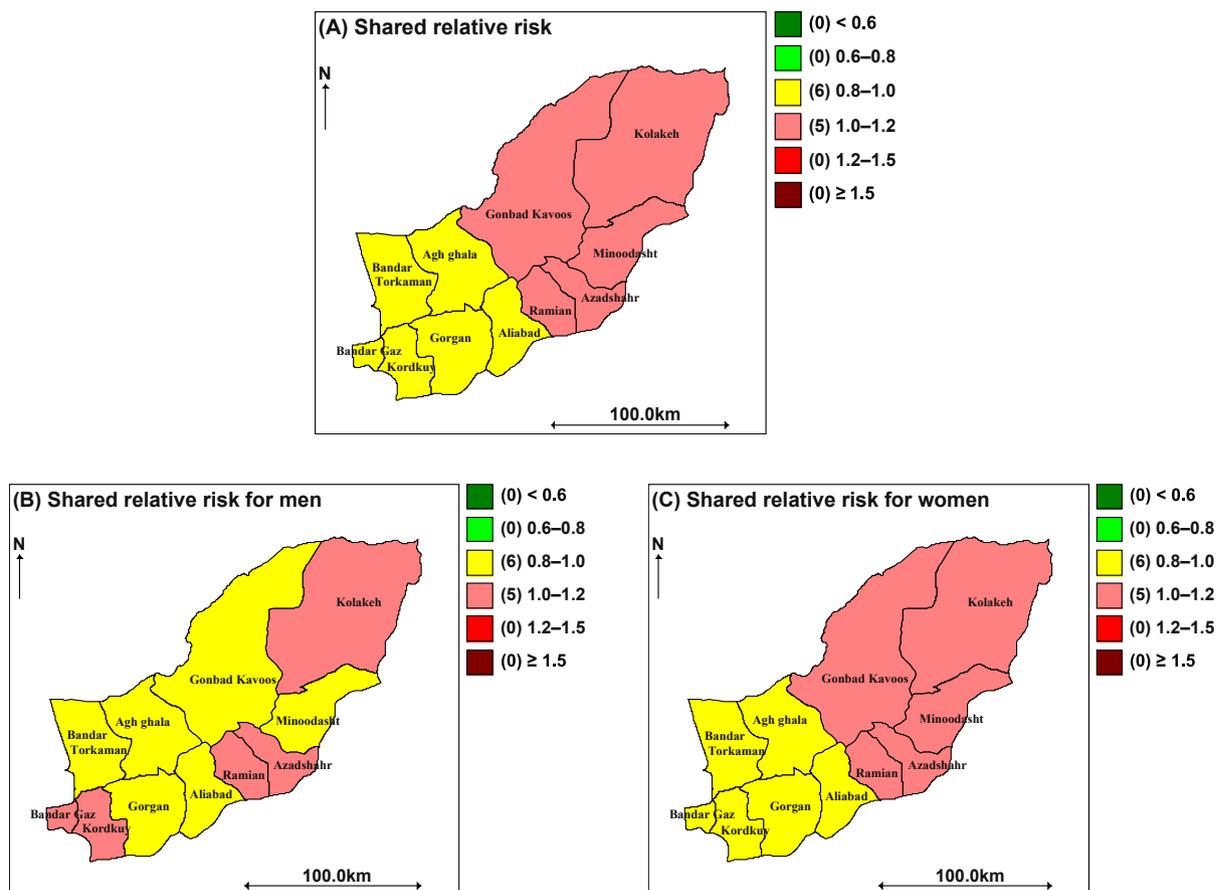

Figure 1. Maps of the posterior median estimated relative risk for two cancers in Golestan, 2004−2008, using a shared component model.
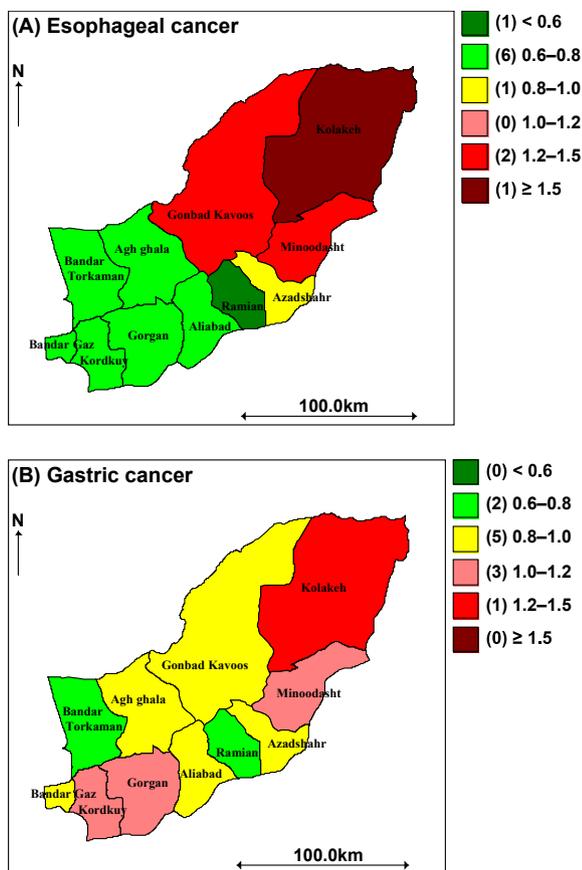
Figure 2. Maps of the posterior median Estimated Relative Risk in the BYM model for two Cancers in Golestan, 2004−2008.

a higher risk than the southern half. Also this pattern remained for women, but for men, the relative risk estimate was distributed across the region.

In addition, we present the individual maps of EC and GC in Figure 2A and B. Figure 2A displays the overall posterior median relative risk surface for EC. Based on this plot, the relative risk of this cancer is higher in the northern part of the area and the concentration of the highest incidence rate is in a northeast county, Kolaleh (>1.5). Furthermore, this map shows that the southern part of the area has a relatively low relative risk (<0.8). Figure 2B presents the pattern of the relative risk estimates from the BYM model for GC, which shows that the cancer incidence risk distributed in total province but the concentration of high incidence is partly in a northeast county (1.2−1.5). Mohebbi et al [3,4] also showed that the northern half of Golestan province was under more high risk than the other part for both cancers.

This type of analysis may be useful for authorities to evaluate the health care system performance and adjust their policies as a result. In our study, the geographical pattern of relative risk using a shared component model indicated that a low fruit and vegetable diet component is important in the target province and more attention is needed in the allocation and delivery of public health policies.

By contrast, although we consider a diet low in fruit and vegetables as a shared component in our study, we can conclude that the other major risk factors, which are common for the two cancers under study, such as low socioeconomic status and tobacco use, should receive more attention in the high-risk areas.

A possible extension to this study would be to include the maps of the incidence rates after adjustment for sex, age, socioeconomic background, etc., or to import a temporal component into the model to improve the correlation more.

The study might have some limitations that caused over-or-under estimation. One of these limitations is the edge effect phenomenon. Although we used the adjacent matrix, some counties in the Golestan province border counties in other regions and the data at hand are limited to the counties under study.

Conflicts of interest

All contributing authors declare no conflicts of interest.

Acknowledgments

References

1. Kolahdoozan S, Sadjadi A, Radmard AR, Khademi H. Five common cancers in Iran. Arch Iran Med 2010 Mar;13(2):143−6.
2. Ramezani Gourabi B. Recognition of geographical diffusion esophagus cancer in southwestern of Caspian Sea. J Am Sci 2011 Feb 25;7(2):297−302.
3. 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.
4. Mohebbi M, Wolfe R, Jolley D, Forbes AB, Mahmoodi M, Burton RC. 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.
5. Mahboubi E, Kmet J, Cook P, Day N, Ghadirian P, Salmasizadeh S. Oesophageal cancer studies in the Caspian Littoral of Iran: the Caspian cancer registry. Br J Cancer 1973 Sep;28(3):197−214.
6. Kamangar F, Malekzadeh R, Dawsey SM, Saidi F. Esophageal cancer in Northeastern Iran: a review. Arch Iran Med 2007 Jan;10(1):70−82.
7. Sadjadi A, Nouraie M, Mohagheghi M, Mousavi-Jarrahi A, Malekezadeh R. Donald Maxwell P. Cancer occurrence in Iran in 2002, an international perspective. Asian Pac J Cancer Prev 2005 Jul−Sep;6(3):359−63.
8. Sadjadi A, Marjani H, Semnani S, Nasseri-Moghaddam S. Esophageal cancer in Iran: a review. Mid E J Cancer 2010;1(1):5−14.

9. Dreassi E. Shared component models in joint disease mapping: a comparison via a simulation experiment. Working Papers Elettronici del Dipartimento di Statistica G. Parenti dell'Universita degli Studi di Firenze; 2010.

10. Lawson AB, Biggeri AB, Boehning D, et al. Disease mapping models: an empirical evaluation. Disease Mapping Collaborative Group. Stat Med 2000 Sep;19(17−18):2217−41.

11. Tzala E, Best N. Bayesian latent variable modelling of multivariate spatio-temporal variation in cancer mortality. Stat Med 2000 Sep 15−30;19(17−18):2217−41.

12. Assunção RM, Castro MS. Multiple cancer sites incidence rates estimation using a multivariate Bayesian model. Int J Epidemiol 2004 Jun;33(3):508−16.

13. Dabney AR, Wakefield JC. Issues in the mapping of two diseases. Stat Methods Med Res 2005 Feb;14(1):83−112.

14. Mahaki B, Mehrabi Y, Kavousi A, et al. Multivariate disease mapping of seven prevalent cancers in Iran using a shared component model. Asian Pac J Cancer Prev 2011;12(9):2353−8.

15. Langford IH, Leyland AH, Rasbash J, Goldstein H. Multilevel modelling of the geographical distributions of diseases. J R Stat Soc Ser C Appl Stat 1999;48(2):253−68.

16. Leyland AH, Langford IH, Rasbash J, Goldstein H. Multivariate spatial models for event data. Stat Med 2000 Sep 15-30;19(17−18):2469−78.

17. Knorr-Held L, Best NG. A shared component model for detecting joint and selective clustering of two diseases. J R Stat Soc Ser A Stat Soc 2001;164(1):73−85.

18. Held L, Natário I, Fenton SE, Rue H, Becker N. Towards joint disease mapping. Stat Methods Med Res 2005 Feb;14(1):61−82.

19. Manda SM, Feltbower RG, Gilthorpe MS. Review and empirical comparison of joint mapping of multiple diseases. S Afr J Epidemiol Infect 2011;27(4):169−82.

20. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math 1991 Mar;43(1):1−20.

21. Downing A, Forman D, Gilthorpe MS, Edwards KL, Manda SO. Joint disease mapping using six cancers in the Yorkshire region of England. Int J Health Geogr 2008 Jul 28;7:41.

22. Richardson S, Abellan JJ, Best N. Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire (UK). Stat Methods Med Res 2006 Aug;15(4):385−407.

23. Ntzoufras I. Bayesian modeling using WinBUGS. Hoboken, NJ: John Wiley & Sons; 2011.