

Research Paper

Comparing Spatial Ecological Regression Models in Breast Cancer Incidence in Iran



Hasti Hashemi¹ , Behzad Mahaki^{2*} , Rahman Farnoosh³

1. Department of Statistics, Science and Research Branch, Islamic Azad University, Tehran, Iran.

2. Department of Biostatistics, School of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran.

3. School of Mathematics, Iran University of Science and Technology, Tehran, Iran.



Citation Hashemi H, Mahaki B, Farnoosh R. Comparing Spatial Ecological Regression Models in Breast Cancer Incidence in Iran. *Journal of Research & Health*. 2024; 14(4):329-340. <http://dx.doi.org/10.32598/JRH.14.4.2014.1>

<http://dx.doi.org/10.32598/JRH.14.4.2014.1>

ABSTRACT

Background: The present study investigates the incidence of breast cancer in Iran at the province level and also explores the impact of the covariates of some risk factors using the temporal risk of cancer, Poisson, log-normal, and Besag-York-Mollie (BYM) models.

Methods: This is an ecological study and the population includes all patients with breast cancer from 2005 and 2009 in Iran. Risk factors for the disease included physical activity and being overweight, according to an annual report on the registration of cancer cases in the country. Individuals with a positive diagnostic test or a biopsy for breast cancer were identified as definitive and registered at the breast cancer registry. Poisson, log-normal, and BYM models were used to estimate the relative risk indices, spatial risk, and temporal risk of cancer.

Results: The total number of registered breast cancers from 2005 to 2009 in Iran was 32 694 cases. Among the provinces, the most observed cases are related to Isfahan Province with 2862 cases, and Khorasan Razavi Province with 2646 cases. The lowest incidence is associated with Kohgiluyeh and Boyer-Ahmad Province with 104 cases. The highest relative risk of breast cancer was observed in 2009 in Tehran Province and the lowest in Kohkiluyeh Boyer-Ahmad Province, and Sistan-Baluchestan Province.

Conclusion: Due to the better fit of the BYM model, it is recommended to use the BYM model in future studies, and considering the significance of the human development index variable, this variable should be considered in similar studies.

Keywords: Disease mapping, Breast cancer, Besag-York-Mollie (BYM) model, Poisson regression model, Normal log model

Article info:

Received: 28 Mar 2022

Accepted: 10 Feb 2024

Publish: 01 Jul 2024

* Corresponding Author:

Behzad Mahaki, Professor.

Address: Department of Biostatistics, School of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran.

E-mail: behzad.mahaki@gmail.com



Copyright © 2024 The Author(s);

This is an open access article distributed under the terms of the Creative Commons Attribution License (CC-BY-NC: <https://creativecommons.org/licenses/by-nc/4.0/legalcode.en>), which permits use, distribution, and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Introduction

Cancer is a disease caused by the abnormal growth of cells in the body. Abnormal growth of such cells eventually leads to the formation of large masses (tumors) [1]. Cancer is one of the leading causes of death worldwide. Despite many efforts to reduce cancer deaths in recent years, cancers are still the second most common cause of death after cardiovascular disease in developed countries and the third leading cause of death in developing countries [2]. According to the [World Health Organization \(WHO\)](#), cancer kills 7.6 (13% of all deaths) millions of people worldwide each year. About 70% of all cancer deaths occur in low and middle-income countries [3].

Estimates show that by 2030, about 21 million new cases of cancer will occur annually, of which 60% to 70% will be in developing countries, and the global burden of cancer is increasing due to aging and population growth, as well as high-risk behaviors, especially smoking. In addition to the psychological and economic consequences, patients with cancer suffer from cancer pain, which adversely affects their quality of life [4].

Today, 45% to 50% of deaths in women aged 45 to 64 years, and 30% of deaths in men aged 45 to 64 years are associated with cancers. This high rate indicates a significant increase in cancer deaths compared to the last century. Air pollution, smoking, diet change, alcohol consumption, stress, and so on can be risk factors for cancer, which has grown significantly in the last century [5].

In Iran, cancer is the third leading cause of death. Every year, more than 30000 people in Iran die of cancer. It is estimated that more than 70000 new cancers occur annually in the country [6].

Cancer rates are expected to rise in the future due to increased life expectancy and modern lifestyle trends [7]. Accordingly, the priority of health policymakers should be to establish a national center for cancer control and prevention. The forecast of health organizations for the incidence of cancer in Iran in 2020 will reach 85653 cases in the total population and the rate of cancer deaths will reach 622897 cases [6].

Of all cancers, breast cancer is the most common and a major cause of cancer death in women in Western countries. The breast is a secretory organ made up of glandular tissue and ducts. The cause of this cancer is unknown; however, due to the rarity of this cancer in men, its etiol-

ogy is due to female hormones [8]. These tumors grow slowly but reach an advanced stage shortly after onset. The variability of the distribution of this cancer in different regions shows the influence of the role of environmental factors in this cancer [9].

According to many studies, physical activity, overweight, and obesity are the most important and measurable factors in the incidence of cancer, so many studies on breast cancer and risk factors have been conducted [10-12]. Few studies have been done at the ecological level, given the geographical distribution [13]. In this study, based on ecological information and mathematical models, changes in breast cancer incidence were considered. Due to the aging population and consequently the upward trend of cancer in Iran and the importance of cancer prevention and the spread of cancer in different parts of the country, it is necessary to identify risk factors and areas with risk factors [14]. In recent years, attention to mapping and the risk of disease has increased significantly because the geographical distribution of incidence, prevalence, and mortality rates has played an important role in identifying risk factors and causes of many diseases and should not be underestimated [15].

A study by Colonna et al. examined how to select and interpret a Bayesian spatial model and a Poisson regression model to explain the variability of small-area cancers. In this study, Besag-York-Mollie (BYM) models were used to map diseases, spatial autocorrelation tests (Moran statistics) of systemic inflammatory response syndrome, disseminated intravascular coagulation criteria to compare different BYM and compare the experimental variance of structural heterogeneity and non-structural heterogeneity of the BYM model [16].

Renart et al. investigated common ecological regression errors of common cancers on the exclusion index. They presented two models of relative risk estimated by the indirect method and the use of systemic inflammatory response syndrome as the response variable (model 1) and the relative risk model estimated using the model, including age as the explanatory variable and crude cancer rate as the response variable (model 2). They compared and found that model 2 fits better while model 1 leads to skewness. Accordingly, if the age variable was considered one of the explanatory variables and the raw rates as the response variable was used to explain the relative risk of cancer outbreak using ecological models that control geographic variability, the raw variable would be less skewed [17].

Hou et al. showed that a healthy eating pattern, which includes eating fruits and vegetables, can reduce the risk of breast cancer. In many studies, researchers analyze data that contains geographic information and provides information about a specific location and space. The data that has such a property is called spatial data [18].

With access to spatial or spatio-temporal information, statistical methods have also been developed to use this data to obtain more accurate information. One of the most important of such methods is disease mapping. Disease mapping is one of the oldest and most important tools for making assumptions about the cause of diseases and identifying areas that need to be studied more closely. The study of spatial (or space-time) changes and the rate of disease is called disease mapping.

Therefore, in the present study, the incidence of breast cancer in the provinces of Iran and the role of risk factors for overweight or obesity and physical activity were investigated using BYM which considers the role of spatial correlations between cancer incidence in the study areas. The models used in this study are complete BYM and experimental Bayesian (Gamma Poisson, log normal) models. Accordingly, the present study investigates the incidence of breast cancer in Iran at the province level and also explores the impact of some covariates of some risk factors by use of the temporal risk of cancer, Poisson, log-normal, and BYM models.

Methods

This study is an ecological analysis that examines the relationship between the prevalence of diseases and risk factors in groups. The regression analysis of this model is based on ecological regression. Due to the non-independence of regions (responses) in geographical studies, to prevent possible bias in estimating regression coefficients, the spatial correlation structure of responses in neighboring regions should be considered. This type of regression is called spatial ecology regression. The Bayesian approach should be used to fit this model. This model is called the Bayesian spatial ecology regression model. In this study, the data from 30 provinces that were registered at the Cancer Registration Center was used.

Information on the incidence of breast cancer was extracted from the annual national reports of cases registered by the Center for Non-Communicable Diseases Management of the Ministry of Health and Medical Education. The data were collected by the Cancer Department and from the cancer registration system. Cancer data were extracted from the Center for Cancer and Non-Communicable Diseases of the Ministry of Health. New cases with definite confirmation of cancer are reg-

istered in the disease registration centers of the province and are referred to the national center annually.

Disease mapping examines the spatial (or space-time) changes in disease rates, and it shows the geographical distribution of a disease within a given population, which determines the spatial pattern of the addresses of several specific diseases. One of the important goals in the analysis of spatial or spatio-temporal data is the use of statistical models to determine the effects of potential risk factors on the occurrence of the desired outcome.

The simplest method of these methods is raw mapping of diseases, which is usually misleading due to the small number of values in the areas and not considering the spatial correlation between areas, in which the relative risk criterion of each area can be used. In that condition, the probability of a person getting an illness in an area is divided by the probability of that person from the population, which is called the standardized mortality rate (SMR).

Statistical methods

Gamma-Poisson models, normal logs, and experimental BYM are the subsets of Bayesian models. In data related to disease mapping, due to lack of data, the use of Bayesian methods that combine data with previous information is a more appropriate method than the SMR method. The following model was used to evaluate the expected incidence of disease in an area and to assess the relationship between incidence and risk factors [19].

In disease mapping, we assume that the study area is divided into a smaller area ($I=1, 2, \dots, I$) and the observed number of deaths due to the disease (or incidence) in the I and E areas. I is the expected number of diseases in that area. Assuming the target community, SMR is defined as follows (Equation 1) [20]:

$$1. SMR_i = \frac{O_i}{E_i} \quad i = 1, 2, \dots, I$$

SMR values are an estimate of the relative risk of each area. To calculate these indices, the map is divided into n adjacent non-interfering regions ($I=1, \dots, n$). The number of observed and expected events in the area I are denoted, respectively, in which it is assumed to be fixed and known in the study period and the product of the population of each area in the total incidence (to interfering regions ($I = 1, \dots, n$)). The number of observed and relative risk of disease in the region I is also indicated in the population in area I (Equation 2).

$$2. \text{var}(SMR_i) = \frac{y_i}{(E_i)^2}$$

In disease mapping, it is assumed that the number of events between regions is independent of each other and follows the Poisson distribution on average.

Poisson distribution

If Y is defined as the number of events of an event in a given spatial or temporal interval such that the average of the event in question in that interval, then Y will have a Poisson distribution with parameter λ (Equation 3) [21]:

$$Y \sim \text{Pois}(\lambda)$$

$$3. P(Y = y) = \frac{e^{-\lambda} \lambda^y}{y!} \quad Y = 0, 1, 2, \dots$$

Bayesian analysis in spatial statistics

In Bayesian inference, parameters are considered as random variables that are used to observe data to update the previous information. At the core of Bayesian analysis is data likelihood. Validation is the co-distribution of the observed data in terms of a parameter or parameter vector (θ). It can also be defined as a function that describes the dependence of the parameters on the sample values. All data information is expressed by the likelihood function. In addition, the probability principle implies that any event that did not occur does not affect the final inference because all inferences are based on the probability and information of the observed data [22].

Normal log model

The normal log model model has limitations because the compatibility of the independent variable is difficult and spatial correlations between regional rates are not possible. The normal log model is more flexible for relative risk (Equation 4).

$$y_i \sim \text{Poisson}(e_i \theta_i)$$

$$4. \log \theta_i = \alpha + v_i$$

$$v_i \sim N(0, \tau_v^2)$$

To examine the effect of prior distribution selection on relative risk estimation, it is important to use sensitive tests for different prior distribution choices. If the data is large then the data overrides the previous distribution. Therefore, in such a case, selecting the initial values of the parameters is less important. If the data is small,

choosing the right combination of initial parameter values becomes important [23].

Results

The summary of the data updated by year and province from 1999 to 2010 are used for analysis and we used several types of data to estimate the mortality of all causes in Iran, including data source system virtual reality (VR) (data source=32), and surveys. We extracted the censuses, from which the mortality call data (data source=9) were summarized, in addition to birth history (5 data sources for SBH and complete birth history) (one data source for CBH). The Ministry of Health and Medical Education of Iran is responsible for providing and managing cancer registration in Iran. Cancer registration information from 2000 to 2010 is available for the whole country. However, the cancer registration system has problems, such as missing data, incomplete cancer registration system, and duplicate data. Meanwhile, the data is only available to some researchers for a limited number of years. Research on non-communicable diseases has been conducted at the University of Tehran.

The total number of registered breast cancers in the years 2005 to 2009 in Iran was 32 694 cases. Among the provinces, the most observed cases are related to Isfahan Province with 2862 cases, and Khorasan Razavi Province with 2646 cases. The lowest incidence is associated with Kohgiluyeh and Boyer-Ahmad provinces with 104 cases. The Table 1 shows the total number of registered and expected breast cancers from 2005 to 2009.

According to the observed and expected values of breast cancer incidence in the provinces, the results showed that at the end of 2009, Tehran Province (n=2088) and Isfahan Province (n=643) had the highest incidence, and South Khorasan Province and Kohkiluyeh and Boyer-Ahmad Province had the lowest cancer rates.

In estimating the proportional risk of breast cancer with and without the involvement of risk factors according to the BYM model, the results showed that the highest incidence of breast cancer was in Tehran Province and the lowest was in Sistan Baluchestan and Kohkiluyeh Boyer-Ahmad provinces (Table 1).

In estimating the relative risk of breast cancer with and without the involvement of risk factors according to the normal log model, the results showed that the highest incidence of breast cancer was in Tehran Province and the lowest was in Sistan Baluchestan and Kohkiluyeh Boyerahmad provinces and the incidence of cancer in

Table 1. Proportional risk of provinces for breast cancer incidence for the Bayesian model

Province	Without Risk Factors					With Risk Factors				
	2005	2006	2007	2008	2009	2005	2006	2007	2008	2009
East Azarbaijan	0.46	0.59	0.74	0.94	1.20	0.46	0.59	0.74	0.94	1.20
West Azarbaijan	0.76	0.69	0.63	0.58	0.53	0.76	0.69	0.63	0.58	0.53
Ardabil	0.52	0.49	0.47	0.45	0.43	0.52	0.49	0.47	0.45	0.43
Isfahan	1.58	1.47	1.37	1.28	1.20	1.57	1.47	1.37	1.28	1.20
Ilam	0.56	0.54	0.53	0.52	0.51	0.52	0.52	0.52	0.52	0.52
Bushehr	0.97	0.96	0.95	0.94	0.93	0.97	0.96	0.95	0.94	0.93
Tehran	1.38	1.41	1.44	1.46	1.49	1.38	1.41	1.44	1.46	1.49
Chaharmahal and Bakhtiari	0.58	0.59	0.60	0.62	0.64	0.55	0.57	0.59	0.62	0.65
South Khorasan	0.54	0.53	0.53	0.53	0.53	0.49	0.50	0.51	0.52	0.54
Khorasan Razavi	1.14	1.09	1.03	0.98	0.93	1.14	1.09	1.03	0.98	0.93
North Khorasan	0.45	0.45	0.45	0.46	0.46	0.44	0.45	0.45	0.46	0.46
Khuzestan	1.07	1.11	1.14	1.18	1.22	1.07	1.11	1.14	1.18	1.22
Zanjan	0.58	0.53	0.49	0.45	0.41	0.59	0.54	0.49	0.45	0.41
Semnan	0.89	0.93	0.97	1.01	1.06	0.91	0.94	0.98	1.01	1.05
Sistan and Baluchestan	0.29	0.26	0.24	0.22	0.21	0.28	0.26	0.24	0.23	0.21
Fars	1.25	1.22	1.20	1.18	1.16	1.24	1.22	1.20	1.18	1.16
Qazvin	0.89	0.85	0.82	0.78	0.75	0.90	0.86	0.82	0.78	0.75
Qom	1.00	0.90	0.80	0.72	0.65	1.04	0.92	0.81	0.72	0.64
Kurdistan	0.60	0.62	0.64	0.66	0.69	0.60	0.62	0.64	0.66	0.69
Kerman	0.44	0.50	0.58	0.66	0.76	0.43	0.50	0.57	0.66	0.76
Kermanshah	0.94	0.93	0.92	0.91	0.89	0.96	0.94	0.92	0.90	0.89
Kohgiluyeh Boyer-Ahmad	0.36	0.36	0.37	0.37	0.37	0.35	0.35	0.36	0.37	0.37
Golestan	0.98	0.97	0.96	0.96	0.95	0.99	0.98	0.97	0.96	0.95
Guilan	1.33	1.23	1.15	1.07	1.00	1.34	1.24	1.15	1.07	1.00
Lorestan	0.74	0.72	0.70	0.68	0.67	0.73	0.71	0.70	0.68	0.67
Mazandaran	1.19	1.18	1.17	1.16	1.15	1.19	1.18	1.17	1.16	1.14
Markazi	0.69	0.77	0.85	0.95	1.05	0.71	0.78	0.86	0.95	1.05
Hormozgan	0.65	0.60	0.56	0.52	0.48	0.65	0.60	0.56	0.52	0.48
Hamedan	0.89	0.86	0.84	0.82	0.80	0.88	0.86	0.84	0.82	0.80
Yazd	1.22	1.22	1.21	1.21	1.20	1.22	1.22	1.21	1.21	1.21

the provinces was lower in the column without the presence of risk factors (Table 2).

In estimating the provincial risk of breast cancer with and without the involvement of risk factors according to the Gamma-Poisson model, the results showed that the highest incidence of breast cancer was in Tehran and Isfahan provinces and the lowest in Kohgiluyeh-Boyer-Ahmad Province and then in Sistan and Baluchestan Province. The results in this study were opposite to other models (Table 3).

According to Table 4, regarding the Gamma-Poisson model, the role of risk factors has become significant. This model does not consider the correlation between provinces. For this reason, using the table related to the Gamma-Poisson model can be misleading. In the log-normal model, considering the non-structural heterogeneity, the effects of the variables were adjusted and the factors of overweight and obesity became significant. The positiveness of this coefficient shows that the increase in overweight and obesity increases the incidence of cancer.

Figure 1A shows the relative risk of provinces without adjusting risk factors and without considering structural and non-structural heterogeneities. According to this map, the provinces of Yazd and Tehran have the highest risk, and Sistan and Baluchestan Province has the lowest risk of breast cancer. Meanwhile, the central provinces are more at risk.

Figure 1B shows the relative risk of the provinces by considering non-structural heterogeneities without modulating the effect of risk factors. According to this map, the northwestern and southeastern provinces have a lower risk of breast cancer, and the provinces of Sistan Baluchestan and Kohgiluyeh and Boyer-Ahmad have the lowest risk of breast cancer.

Figure 1C shows the relative risk of the provinces by considering non-structural heterogeneity and by modifying the risk factors. According to this map, Khorasan Razavi and Hamedan provinces have the highest risk of breast cancer, and Kohgiluyeh and Boyer-Ahmad, Sistan and Baluchestan, and Ardabil provinces have the lowest risk of breast cancer.

Figure 1D shows the relative risk of the provinces without modulating the effect of risk factors and taking into account structural and non-structural heterogeneities. According to this map, Isfahan and Tehran provinces

have the highest risk of Sistan and Baluchestan, Kohgiluyeh, and Boyer-Ahmad provinces.

Figure 1E shows the relative risk of the provinces by modulating the effect of risk factors and considering structural and non-structural heterogeneities. According to this map, the provinces of Yazd, Qazvin, Ardabil, and North Khorasan have the lowest risk of breast cancer, and the provinces of Khorasan Razavi, Khorramabad, and Hamedan have the highest risk.

Discussion

According to the BYM model, without adjusting the effect of risk factors, the provinces of Isfahan, Yazd, and Tehran have the highest risk of breast cancer, followed by the provinces of North, Fars, Khuzestan, and North Khorasan, and the northeastern and southwestern provinces have the highest risk. Among the provinces, Sistan and Baluchistan Province and Chaharmahal and Bakhtiari Province had the lowest risk of infection. These results are consistent with the results of the study by Khoshkar et al [24]. In the present study, the time trend of cancer incidence was estimated in 27% of the rising regions. Importantly, the estimation of the rising trend of risk for provinces with low cancer risk, which indicates changes in the pattern of cancer incidence in these provinces and the need for serious interventions. Meanwhile, the great distance of most of these areas from the center of the country and their location in the border points, which are economically different from other provinces of the country is considerable. Many studies have found the role of distance and economic status to be effective in controlling and reducing the incidence and complications of cancer.

In the log-normal model, considering the non-structural heterogeneity, the effects of the variables were adjusted and the factors of overweight and obesity became significant. The positiveness of this coefficient shows that the increase in overweight and obesity increases the incidence of cancer. The log-normal model does not consider the structural correlation between provinces. The most complete model is the BYM model, which considers both structural and non-structural correlations. Considering the structural and non-structural heterogeneity, none of the risk factors are significant. Table 4 shows the comparison of the goodness of fit of Gamma-Poisson, log-normal, and BYM models with and without risk factors using the deviance information criterion (DIC) index. The BYM model has the best fit without the presence of risk factors because it includes structural and non-structural heterogeneities. In this model, the effect of risk factors is moderated by considering structural

Table 2. Relative risk of provinces for breast cancer incidence for the normal log model

Provinces	Without Risk Factors					With Risk Factors				
	2005	2006	2007	2008	2009	2005	2006	2007	2008	2009
East Azarbaijan	0.4625	0.5863	0.7436	0.9435	1.198	0.4637	0.5874	0.7444	0.9439	1.197
West Azarbaijan	0.7593	0.694	0.6346	0.5807	0.5317	0.7583	0.6933	0.6342	0.5805	0.5316
Ardabil	0.5234	0.4958	0.4703	0.4468	0.4251	0.5225	0.4953	0.4702	0.447	0.4257
Isfahan	1.578	1.472	1.373	1.281	1.196	1.575	1.47	1.373	1.282	1.197
Ilam	0.5632	0.5469	0.5323	0.5191	0.5074	0.527	0.5229	0.5199	0.518	0.5172
Bushehr	0.9723	0.959	0.9466	0.9353	0.9249	0.9776	0.963	0.9496	0.9371	0.9256
Tehran	1.383	1.41	1.437	1.464	1.493	1.383	1.409	1.436	1.464	1.493
Chaharmahal and Bakhtiari	0.5727	0.5871	0.6027	0.6196	0.6378	0.5477	0.5697	0.5934	0.6189	0.6464
Southern Khorasan	0.5381	0.5334	0.5298	0.5273	0.5257	0.4952	0.504	0.5139	0.5251	0.5376
Khorasan Razavi	1.142	1.085	1.03	0.9786	0.9298	1.143	1.085	1.03	0.9786	0.9297
North Khorasan	0.4475	0.4499	0.453	0.4569	0.4616	0.4411	0.4455	0.4508	0.4568	0.4636
Khuzestan	1.072	1.107	1.144	1.181	1.221	1.068	1.104	1.142	1.182	1.223
Zanjan	0.5912	0.5369	0.4885	0.4452	0.4064	0.5937	0.5386	0.4894	0.4455	0.4062
Semnan	0.8835	0.9224	0.9643	1.009	1.058	0.8965	0.9316	0.9693	1.01	1.053
Sistan and Baluchestan	0.2896	0.2658	0.2442	0.2248	0.2072	0.2872	0.2643	0.2436	0.2249	0.2079
Fars	1.244	1.222	1.2	1.179	1.159	1.242	1.22	1.199	1.179	1.159
Qazvin	0.8809	0.8464	0.814	0.7836	0.755	0.8948	0.8549	0.8175	0.7826	0.7498
Qom	0.9986	0.8954	0.8038	0.7223	0.6499	1.033	0.9147	0.8111	0.7201	0.6399
Kurdistan	0.5979	0.6184	0.6402	0.6634	0.6881	0.5926	0.6147	0.6382	0.6633	0.6901
Kerman	0.437	0.501	0.5748	0.66	0.7583	0.4279	0.4941	0.571	0.6602	0.764
Kermanshah	0.9481	0.933	0.9187	0.9051	0.8923	0.9607	0.9412	0.9226	0.9049	0.8882
Kohgiluyeh Boyer-Ahmad	0.3622	0.3633	0.3652	0.3678	0.3711	0.3462	0.3518	0.3583	0.3655	0.3737
Golestan	0.9839	0.9742	0.9652	0.9569	0.9492	0.9937	0.9808	0.9687	0.9574	0.9468
Guilan	1.326	1.235	1.15	1.071	0.998	1.335	1.24	1.152	1.071	0.9955
Lorestan	0.7349	0.7167	0.6995	0.6833	0.668	0.7335	0.7159	0.6993	0.6836	0.6688
Mazandaran	1.188	1.176	1.166	1.155	1.146	1.195	1.181	1.168	1.155	1.143
Markazi	0.6906	0.7663	0.851	0.9459	1.052	0.7013	0.7743	0.8556	0.9463	1.048
Hormozgan	0.6446	0.5993	0.5578	0.5198	0.4848	0.6512	0.6034	0.5597	0.5197	0.4831
Hamedan	0.8822	0.8592	0.8374	0.8167	0.797	0.8811	0.8584	0.8368	0.8162	0.7967
Yazd	1.237	1.226	1.216	1.207	1.199	1.238	1.227	1.217	1.208	1.2



Table 3. Provincial risk for breast cancer incidence for the Gamma-Poisson model

Provinces	Without Risk Factors					With Risk Factors				
	2005	2006	2007	2008	2009	2005	2006	2007	2008	2009
East Azarbaijan	1.012	0.9868	0.962	0.9379	0.9145	0.7665	0.8234	0.8846	0.9506	1.022
West Azarbaijan	0.8928	0.7675	0.66	0.5675	0.4881	0.818	0.7265	0.6453	0.5734	0.5096
Ardabil	0.8265	0.658	0.524	0.4175	0.3327	0.7234	0.6078	0.5108	0.4296	0.3614
Isfahan	1.1	1.164	1.233	1.305	1.382	1.134	1.188	1.245	1.305	1.368
Ilam	0.8578	0.7089	0.5863	0.4852	0.4018	0.4052	0.4329	0.4629	0.4955	0.5308
Bushehr	1.007	0.9771	0.948	0.92	0.8931	1.105	1.044	0.9861	0.9319	0.8809
Tehran	1.133	1.236	1.349	1.472	1.606	1.247	1.317	1.391	1.47	1.552
Chaharmahal and Bakhtiari	0.9093	0.7964	0.6977	0.6115	0.5361	0.6195	0.6189	0.6187	0.6189	0.6195
Southern Khorasan	0.8668	0.7238	0.6047	0.5055	0.4228	0.4038	0.4369	0.4732	0.513	0.5565
Khorasan Razavi	1.025	1.012	0.999	0.9863	0.9737	0.9809	0.9831	0.9855	0.9881	0.9907
North Khorasan	0.8363	0.6737	0.543	0.4379	0.3533	0.6693	0.5856	0.5127	0.4491	0.3936
Khuzestan	1.073	1.108	1.144	1.182	1.221	0.9374	1.014	1.096	1.185	1.282
Zanjan	0.8238	0.6538	0.519	0.4123	0.3276	0.7823	0.6368	0.5185	0.4224	0.3443
Semnan	1.031	1.024	1.017	1.011	1.005	1.21	1.139	1.072	1.01	0.9522
Sistan Baluchestan	0.68	0.4453	0.2918	0.1912	0.1254	0.5394	0.3906	0.2831	0.2054	0.1491
Fars	1.074	1.11	1.147	1.186	1.227	1.108	1.133	1.158	1.184	1.211
Qazvin	0.9665	0.8997	0.8377	0.78	0.7266	1.067	0.9592	0.8621	0.775	0.6969
Qom	0.9482	0.8658	0.7909	0.7226	0.6603	1.225	1.018	0.846	0.7037	0.5856
Kurdistan	0.9208	0.8165	0.7242	0.6424	0.57	0.7172	0.6975	0.6786	0.6604	0.6429
Kerman	0.9226	0.8197	0.7284	0.6473	0.5753	0.6132	0.6298	0.647	0.6649	0.6836
Kermanshah	1.001	0.9655	0.9311	0.898	0.8662	1.367	1.177	1.013	0.8724	0.7515
Kohgiluyeh Boyer-ahmad	0.7946	0.6083	0.466	0.3573	0.2741	0.5039	0.4537	0.4089	0.3688	0.333
Golestan	1.015	0.9916	0.9691	0.9472	0.9259	1.176	1.093	1.015	0.943	0.8764
Guilan	1.049	1.061	1.072	1.083	1.095	1.31	1.224	1.144	1.07	1.001
Lorestan	0.9305	0.8339	0.7474	0.67	0.6008	0.6289	0.6466	0.6649	0.684	0.704
Mazandaran	1.068	1.098	1.129	1.161	1.195	1.353	1.28	1.212	1.147	1.086
Markazi	1.013	0.9873	0.9629	0.9392	0.9162	1.301	1.159	1.033	0.9216	0.8222
Hormozgan	0.866	0.7222	0.6025	0.5027	0.4196	0.9463	0.7657	0.6199	0.502	0.4068
Hamedan	0.9758	0.9171	0.8619	0.8102	0.7617	0.6808	0.7236	0.7694	0.8183	0.8705
Yazd	1.075	1.112	1.151	1.191	1.234	1.233	1.219	1.206	1.194	1.182



Table 4. Estimation of the effect coefficients of breast cancer risk factors for ecological regression models

Risk Factors	Gamma-Poisson		Log-normal		Bayesian Model	
	Coefficient	95 % Confidence Interval	Coefficient	95 % Confidence Interval	Coefficient	95 % Confidence Interval
Intercept	-4.1350	(4.4620, -3.8150)	-5.5720	(-7.3020, -4.0480)	-5.4550	(-7.6180, -3.7530)
Overweight and obesity	-0.0171	(-0.0243, 0.0127)	0.0225	(0.0010, 0.0436)	0.0227	(-0.0019, 0.0436)
Physical activity	-0.0043	(-0.00803, -0.0021)	0.00368	(-0.01100, 0.0180)	0.00315	(-0.0110, 0.0196)
DIC for a model with risk factors	887.4000		296.3000		268.2000	
DIC for the model without risk factors	-		297.1000		169.3000	



and non-structural heterogeneities, and there is no need to consider risk factors. The Gamma-Poisson model has the worst fit without the presence of risk factors. Considering the risk factors, the BYM and log-normal models have almost the same fit. In the log-normal model, the inclusion of risk factors did not change the goodness of fit of the model. The Gamma-Poisson model has a poor fit because this model does not consider spatial correlations between provinces. Therefore, it is better to use the BYM model among ecological regression models in ecological analysis.

Conclusion

The unadjusted BYM model had the best fit among the considered models. Without adjusting the effect of risk factors, the provinces of Isfahan, Yazd, and Tehran have the highest incidence of breast cancer and the provinces of Sistan and Baluchistan, Chaharmahal and Bakhtiari have the lowest incidence. By adjusting the risk factors, Khorasan-Razavi, Lorestan, and Hamedan provinces have the highest, and Ardabil, Kohgiluyeh and Boyer Ahmad provinces have the lowest relative risk. For pros-

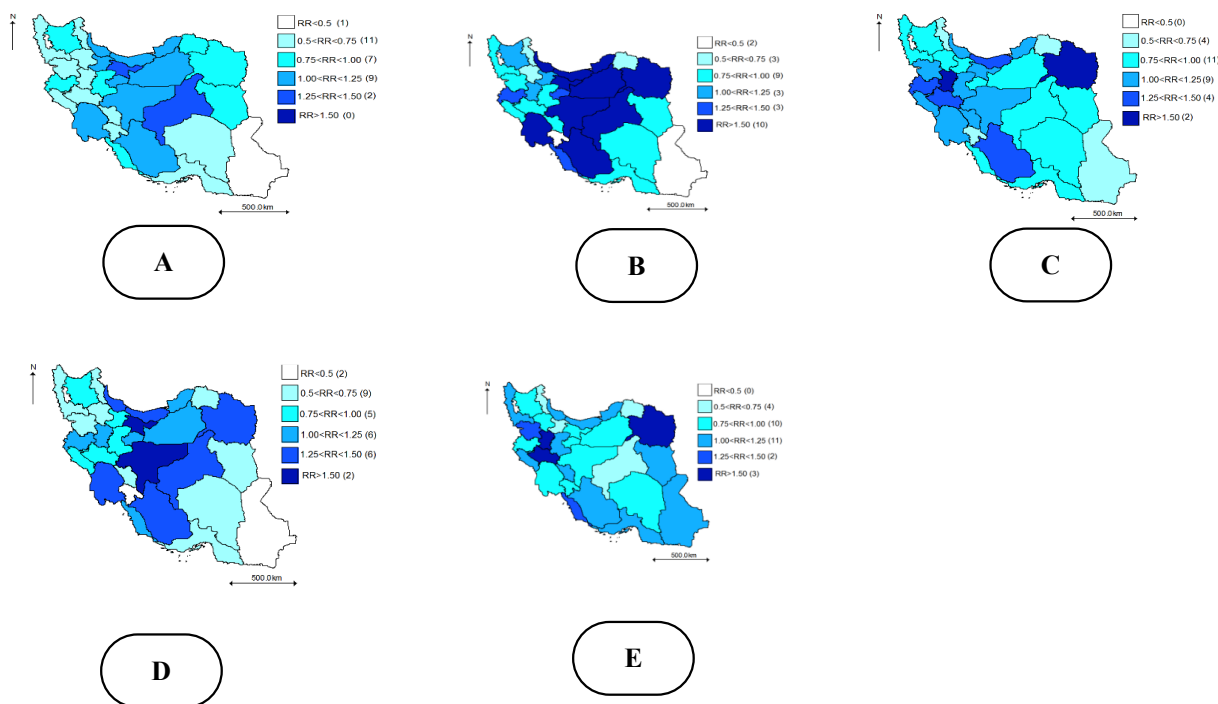


Figure 1. Maps of provincial risk for breast cancer



A) Gamma-Poisson model without modulating the effect of risk factors, B) Proportional risk of provinces for incidence of log-normal breast cancer without modifying the effect of risk factors, C) Log-normal model by modulating the effect of risk factors, D) The relative risk for breast cancer for the Bayesian model without modifying the effect of risk factors, E) Bayesian model breast cancer incidence by modulating risk factors

tate cancer in the unadjusted model, Fars, Semnan, Isfahan, and Tehran provinces have the highest, and Sistan and Baluchistan province has the lowest relative risk. By adjusting the effect of risk factors, Fars and Zanjan provinces have the highest relative risk and Kerman, North Khorasan, Kohgiluyeh and Boyer Ahmad, Qazvin, and Kermanshah provinces have the lowest relative risk.

Study limitations

Air pollution, family history, neonate feeding situation, and other covariates were not available at the province level. So, we suggest conducting further ecological research regarding these factors as well.

Ethical Considerations

Compliance with ethical guidelines

In this study, the principle of confidentiality of information was observed and no information except for the conditions of the research was placed in the possession of any organization or organization.

Funding

This study was funded by [Islamic Azad University, Science and Research Branch](#).

Authors' contributions

Study design: Hasti Hashemi; Methodology: Behzad Mahaki; Data collection: Hasti Hashemi and Rahman Farnoosh; Data analysis: Hasti Hashemi and Behzad Mahaki.

Conflict of interest

The authors declared no conflict of interest.

Acknowledgments

The authors are thankful to all researchers and breast cancer specialists who kindly participated in this study.

References

- [1] Bovero A, Gottardo F, Botto R, Tosi C, Selvatico M, Torta R. Definition of a good death, attitudes toward death, and feelings of interconnectedness among people taking care of terminally ill patients with cancer: An exploratory study. *The American Journal of Hospice & Palliative Care*. 2020; 37(5):343-9. [DOI:10.1177/1049909119883835] [PMID]
- [2] Yang B, Choi H, Lee SK, Chung SJ, Yeo Y, Shin YM, et al. Risk of coronavirus disease 2019 occurrence, severe presentation, and mortality in patients with lung cancer. *Cancer Research and Treatment*. 2021; 53(3):678-84. [DOI:10.4143/crt.2020.1242] [PMID]
- [3] WHO. WHO report on cancer: Setting priorities, investing wisely and providing care for all. Geneva: 2020. [Link]
- [4] Williams F, Zoellner N, Hovmand PS. Understanding Global Cancer Disparities: The role of social determinants from system dynamics perspective. *Transdisciplinary Journal of Engineering & Science*. 2016; 7:10.22545/2016/00072. [DOI:10.22545/2016/00072] [PMID]
- [5] Curtin SC. Trends in cancer and heart disease death rates among adults aged 45-64: United States, 1999-2017. *National vital Statistics Reports*. 2019; 68(5):1-9. [PMID]
- [6] Farhood B, Geraily G, Alizadeh A. Incidence and mortality of various cancers in Iran and compare to other countries: A review article. *Iranian Journal of Public Health*. 2018; 47(3):309-16. [PMID]
- [7] You W, Henneberg M. Cancer incidence increasing globally: The role of relaxed natural selection. *Evolutionary Applications*. 2018; 11(2):140-52. [DOI:10.1111/eva.12523] [PMID]
- [8] Ghoncheh M, Pournamdar Z, Salehiniya H. Incidence and mortality and epidemiology of breast cancer in the world. *Asian Pacific Journal of Cancer Prevention*. 2016; 17(S3):43-6. [DOI:10.7314/APJCP.2016.17.S3.43] [PMID]
- [9] Frugtinet B, Jiang WG, Martin TA. Role of the WASP and WAVE family proteins in breast cancer invasion and metastasis. *Breast Cancer (Dove Med Press)*. 2015; 7:99-109. [DOI:10.2147/BCTT.S59006] [PMID]
- [10] Ballard-Barbash R, Hunsberger S, Alciati MH, Blair SN, Goodwin PJ, McTiernan A, et al. Physical activity, weight control, and breast cancer risk and survival: Clinical trial rationale and design considerations. *Journal of the National Cancer Institute*. 2009; 101(9):630-43. [DOI:10.1093/jnci/djp068] [PMID]
- [11] Leray H, Malloizel-Delaunay J, Lusque A, Chantalat E, Bouglon L, Chollet C, et al. Body Mass Index as a major risk factor for severe breast cancer-related lymphedema. *Lymphatic Research and Biology*. 2020; 18(6):510-6. [DOI:10.1089/lrb.2019.0009] [PMID]
- [12] Rouanet P, Roger P, Rousseau E, Thibault S, Romieu G, Mathieu A, et al. HER2 overexpression a major risk factor for recurrence in pT1a-bN0M0 breast cancer: Results from a French regional cohort. *Cancer Medicine*. 2014; 3(1):134-42. [DOI:10.1002/cam4.167] [PMID]
- [13] Kitaeva AB, Gorshkov AP, Kirichek EA, Kusakin PG, Tsyganova AV, Tsyganov VE. General patterns and species-specific differences in the organization of the tubulin cytoskeleton in indeterminate nodules of three legumes. *Cells*. 2021; 10(5):1012. [DOI:10.3390/cells10051012] [PMID]

- [14] Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, et al. Risk factors and preventions of breast cancer. *International Journal of Biological Sciences*. 2017; 13(11):1387-97. [DOI:10.7150/ijbs.21635] [PMID]
- [15] Oertelt-Prigione S, Seeland U, Kendel F, Rucke M, Flöel A, Gaissmaier W, et al. Cardiovascular risk factor distribution and subjective risk estimation in urban women--the BEFRI study: A randomized cross-sectional study. *BMC Medicine*. 2015; 13:52. [DOI:10.1186/s12916-015-0304-9] [PMID]
- [16] Colonna M, Sauleau EA. How to interpret and choose a Bayesian spatial model and a Poisson regression model in the context of describing small area cancer risks variations. *Revue D'epidemiologie et de Sante Publique*. 2013; 61(6):559-67. [DOI:10.1016/j.respe.2013.07.686] [PMID]
- [17] Renart G, Saez M, Saurina C, Marcos-Gragera R, Ocaña-Riola R, Martos C, et al. A common error in the ecological regression of cancer incidence on the deprivation index. *Revista Panamericana de Salud Publica=Pan American Journal of Public Health*. 2013; 34(2):83-91. [PMID]
- [18] Hou R, Wei J, Hu Y, Zhang X, Sun X, Chandrasekar EK, et al. Healthy dietary patterns and risk and survival of breast cancer: A meta-analysis of cohort studies. *Cancer Causes & Control*. 2019; 30(8):835-46. [DOI:10.1007/s10552-019-01193-z] [PMID]
- [19] Thygesen HH, Zwinderman AH. Modeling Sage data with a truncated gamma-Poisson model. *BMC Bioinformatics*. 2006; 7:157. [DOI:10.1186/1471-2105-7-157] [PMID]
- [20] Yanagimoto T, Kashiwagi N. Empirical Bayes methods for smoothing data and for simultaneous estimation of many parameters. *Environmental Health Perspectives*. 1990; 87:109-14. [DOI:10.1289/ehp.9087109] [PMID]
- [21] Joe H, Zhu R. Generalized Poisson distribution: The property of mixture of Poisson and comparison with negative binomial distribution. *Biometrical Journal*. 2005; 47(2):219-29. [DOI:10.1002/bimj.200410102] [PMID]
- [22] Wright DK, MacEachern S, Lee J. Analysis of feature intervisibility and cumulative visibility using GIS, Bayesian and spatial statistics: A study from the Mandara Mountains, northern Cameroon. *PLoS One*. 2014; 9(11):e112191. [DOI:10.1371/journal.pone.0112191] [PMID]
- [23] Gajewski BJ, Sedwick JD, Antonelli PJ. A log-normal distribution model of the effect of bacteria and ear fenestration on hearing loss: A Bayesian approach. *Statistics in Medicine*. 2004; 23(3):493-508. [DOI:10.1002/sim.1606] [PMID]
- [24] Khoshkar AH, Koshki TJ, Mahaki B. Comparison of bayesian spatial ecological regression models for investigating the incidence of breast cancer in Iran, 2005- 2008. *Asian Pacific Journal of Cancer Prevention*. 2015;16(14):5669-73. [DOI: 10.7314/apjcp.2015.16.14.5669] [PMID]

This Page Intentionally Left Blank