Geographical Spread of Colon and Esophagus Cancers Incidence, in Libya using a Multivariate Spatial Model

Maryam Ahmed Alramah
Department of Statistics, Faculty of Science, University of Tripoli, Tripoli, Libya

الملخص

اقترحت الأبحاث الحالية أنماط الحوادث (الاصابات) وعامل خطر زيادة الوزن المرتبط بسرطان القولون والمريء. تم تحديد معدل حدوث هذه السرطانات باستخدام العديد من الطرق مثل الطبقية حسب الجنس أو بشكل عام. يوضح هذا البحث التباين الجغرافي في حدوث هذين النوعين من السرطانات ذات الصلة للتحقيق في الأهمية النسبية لعامل خطر زيادة الوزن في ليبيا. تم الحصول على بيانات الإصابة بسرطان القولون وسرطان المريء بين عامي 2015 و 2020 من معهد الأورام الأفريقي ، صبراتة ، ليبيا. تم تسجيل هذه البيانات كملاحظات جديدة لـ 22 منطقة سنويًا. تم تطبيق نموذج بايزي متعدد المتغيرات أو نموذج مكون مشترك بايزي للعديد من الأمراض لتحليل التباين المكاني لمعدلات الإصابة بشكل مشترك. في هذا البحث, يظهر التحليل أن الطريقة المشتركة تعطي تقديرات أفضل للمخاطر النسبية مقارنة باستخدام نموذج BYM ونموذج الخليط. بين عامي 2015 و 2020 ، كانت معدلات الإصابة المشتركة بالسرطان مرتفعًا نسبيًا (0.5- 1.5) في ليبيا. أوضحت الخريطة المتعددة أن الأجزاء الغربية والجنوبية من البلاد كانت أكثر عرضة للخطر من الأجزاء الأخرى. المكون الذي يمثل زيادة الوزن كان له تأثير متوسط لسرطان القولون والمريء. تم الحصول على نمط خطر الإصابة هذا للرجال ولكن ذلك كان مختلفاً قليلاً بالنسبة للنساء. يؤدي استخدام نموذج مكون مشترك للنمذجة المشتركة لمعدلات الإصابة إلى تقديرات أكثر دقة من نماذج المرض الفردي. عامل الخطر المشترك (أي زيادة الوزن) مهم ويحتاج إلى مزيد من الاهتمام في تخصيص وتنفيذ سياسات الصحة العامة.

الكلمات المفتاحية: بيانات مرض السرطان، ليبيا، نموذج المكون المشترك ، نموذج BYM ، نموذج الخليط

Abstract

Current researches proposed the incident patterns and the related overweight risk factor for colon and oesophagus cancers. The incidence of these cancers was mapped using many methods such as stratified by sex or in general. This research models the geographical variation in the incidence of these two related cancers to investigate the relative importance of an overweight risk factor in Libya. Data on the incidence of colon cancer and oesophagus cancer between 2015 and 2020 were obtained from the African Oncology Institute, Sabratha, Libya. These data were recorded as new observations in 22 districts yearly. The Bayesian multivariate model or Bayesian shared component model for several diseases was applied to analyse the spatial variation of incidence rates jointly. In this research. the analysis shows that the joint method gives a better relative risk estimates compared with using the BYM model and mixture model. Between 2015 to 2020, the joint incidence rates of the two cancers were relatively high (0.5-1.5) in Libya. The multiple map explained that the western and southern parts of the country were at higher risk than other parts. The component representing overweigh had medium effect of colon and esophagus cancers. This incidence risk pattern has been obtained for the men but that for women have been a little different. Using a shared component model for joint modelling of incidence rates leads to more precise estimates than models of individual disease. A common risk factor (i.e., overweight) is important and needs more attention in the allocation and delivery of public health policies.

Keywords: Libya, Cancer disease data, Shared component mode, BYM model, Mixture model.

Introduction

Cancer is one of the main causes of death and nearly 4,000 new cases of cancer occur annually in Libya. This disease has spread dangerously. Several cancers are related to the gastrointestinal and digestive cancers. In women, the two important cancers are breast and colon: in men, prostate and oesophagus. After lung cancer, these cancers are the main cancers in Libya and the leading cause of deaths (Sabratha Cancer Registry, 2008; Afaf Aburwais et al., 2021). There is evidence of sharp gradients in incidence rates of colon and oesophagus cancer over proportionally wide geographical distances in Libya (El Mistiri et al., 2010; Alramah et al., 2019).

In Libya, cancer is one of causes of death after heart disease, because these cancers had a striking incidence (Sabratha Cancer Registry, 2008). Several researches have highlighted a positive correlation between standardised

incidence ratios of colon and oesophagus cancer which may be evidence of shared common risk factors such as overweight or obesity, smoking, alcohol, and low socio-economic status, but in Libya, the first two components were more effective (Libyan National Statistics Figures, 2011).

In the field of epidemiology, disease mapping has long been applied in the statistical analysis of geographical variation of disease rate (Dreassi et al., 2010), because it gives us beneficial information such as assessment hypotheses, describing areas of unusually high risk, and producing a clean map of disease risk to allocate better resources and public health policies (Lawson et al., 2000).

Disease risk mapping to estimate relative risk based on the most common statistics used in disease mapping is called standardised mortality ratio or standardised morbidity ratio (SMR). It is defined as the ratio of observed to expected count in the region under study and specifies the geographic dispersion of disease morbidity and mortality rates (Tzala et al., 2000). Although this method gets unbiased estimators of relative risk, it suffers from certain drawbacks such as being based on a ratio estimator and the mean and variance of SMR are highly dependent on the expected number of incident cases. It is very large in areas where the expected numbers of cases are small, and small for areas where the expected numbers of cases are large. Furthermore, in areas where there are no observed count data or cases, the SMR is necessarily zero (Lawson et al., 2003).

A variety of alternative models has been suggested to address these drawbacks. Among them, the Bayesian approach is proposed because of its great flexibility in modelling options and a reliable output for inferential purposes. It supposes spatial correlation of disease rates among neighbouring regions to capture the geographical structure, thereby making the estimates of the parameters in the model are more factual (Tzala et al., 2000).

Assucão et al. (2004) showed that most studies in geographical modelling of diseases are focused on only one disease (a single disease). Joint disease mapping or multivariate disease mapping was introduced because several diseases have common risk factors. Many researchers defined joint disease mapping as the spatial modelling of two or more diseases or the same disease in two or more subsets of the population at risk (Tzala et al., 2000; Dabney et al., 2005).

Multivariate modelling of many diseases improves the precision of estimation of underlying disease patterns. Moreover, when there is interest in a relatively rare disease, the ability to assess shared and specific geographic patterns of risk

among different diseases strengthens the relevant results of the rare disease (Dabney et al., 2005).

In recent years, many methods have been suggested for multivariate disease mapping (Downing et al., 2008). Langford et al. (1999) and Leyland et al. (2000) used a multilevel model and first introduced joint spatial model analysis. Knorr-Held and Best (2001) suggested a shared component model, and Held (2005) extended a shared component model to analyse the spatial variation of many diseases that allows the linear predictor to be decomposed into shared and disease specific spatial variability components. Dabney et al. (2005) used multiple modelling of two diseases using a proportional mortality model. Manda et al. (2011) used four joint modelling techniques to compare between them. These are multivariate multiple membership multiple classification models, multivariate intrinsic conditional autoregressive model, and proportional mortality models using data for two cancers (oesophagus and gastric) and shared component model. This study confirmed that the shared component model adds more versatility in answering more substantive epidemiological questions than the other three models.

In Libya, there are no studies that consider the estimation of relative risk for cancer using the Bayesian approach. Therefore, this study applies a shared component model for joint modelling of colon and oesophagus cancers, for which overweight is considered a major risk factor, to explore the geographical variation in incidence rates of these two diseases. In addition, we explore the differences of incidence rates between females and males by joint modelling by gender.

2. Materials and Methods

Data on cases of colon (C18 code) and oesophagus (C15 code) cancers for the six years from 2015 to 2020 were extracted from African Oncology Institute (AOI), Sabratha, in 22 districts of Libya. Relative risk (RR) for each cancer site were calculated (with the number of expected cases calculated using the average number of cases per ward observed in Libya and the population in the 2011 report, because this year is considered a year of political stability).

In this study, assume that O_{ij} presents that observed count for disease j in region i. and E_{ij} indicates the expected number of cases, which calculated by multiplying the overall incidence rate and the estimate of the ward population. The observed count follows Poisson distribution with variance and means $\mu_{ij} = \theta_{ij} * E_{ij}$ in which θ_{ij} is the unknown parameter (relative risk R_{ij}) in the

model. The maximum likelihood estimate of the incidence rate is got by dividing the observed count to expected count for cancer j in area i, as follows: $\theta_{ij} = O_{ij} / E_{ij}$. As earlier mentioned in the introduction, this estimation has some problems. To address these drawbacks, we use the BYM model and mixture model. These models give more reliable estimates for relative risk by obtaining information from neighbouring regions (Besag et al., 1991; Lawson and Clark, 2002). Three models are considered in this study and the construction method for each model is elaborated in the following sections.

Besag, York and Mollie (BYM) Model

To address the problem of SMRs, in this research the Besag, York and Mollie (BYM) model will be used to demonstrate the data, as well as to consider the information of the adjacent neighbours of each district. In this model, there are two sources of changes for explaining the heterogeneity the rate of incidence in every area in addition to independent variables. By other words, the main idea for this model is to produce a more reliable estimation for relative risks and for small areas or rare disease. This is by borrowing required information from the neighbouring areas. Therefore, the model introduced by Clayton and Kaldor (1987) and developed by Besag et al., (1991), is formulated as follows:

$$Log(\theta_{ij}) = \alpha + u_{ij} + v_{ij}$$

Where $\theta_{ij} = \exp(\alpha_j + u_{ij} + v_{ij})$ and α is an overall level of the relative risk. In this model, the log of disease specific area-level relative risks are decomposed into the sum of two components:

1. The first component is u_i that takes into account the effects that vary in a structured manner in space (clustering or correlated heterogeneity). In other words, it is structured heterogeneity (spatial clustering or spatial autocorrelation). It is assuming weights for adjacent areas. For the first component which is the clustering component, a spatial correlation structure is used, where the estimation of the risk in any area depends on neighbouring areas. The conditional autoregressive (CAR) model (Besag et al., 1991) is used

$$[u_i / u_j, i \neq j, \tau_u^2] \sim N(\bar{u}_i, \tau_i^2)$$

$$\bar{u}_i = \frac{\sum_j u_i \omega_{ij}}{\sum_j \omega_{ij}}, \quad \tau_i^2 = \frac{\tau_u^2}{\sum_j \omega_{ij}}; \quad \omega_{ij} = \begin{cases} 1 & if \quad i, j \, are adjectent \\ 0 & if \quad i, j, are not adjectent \end{cases}$$

In other words, the conditional distribution of each area-specific spatially structured component if follow a normal distribution with mean and variance (the average of its neighbours, inversely proportional to the number of its neighbours).

2. The second component is v_i that takes into account the effects that vary in an unstructured way between locations (uncorrelated heterogeneity). The uncorrelated heterogeneities are assumed to follow a normal distribution zero mean and variance of τ_v^2, as follows: $v_{ij} \sim N(0, \tau_v^2)$.

Both τ_u^2 and τ_v^2 parameters control variability of u and v. To analysis of full Bayesian, the prior distributions should be determined for these parameters. Researchers have proposed Gamma prior distributions for these parameters (Lawson et al., 2003; Bernardinelli et al., 1995; Besag et al., 1991).

Mixture Model

The mixture model proposed by Lawson and Clark (2002) will be considered in this research. This model allows the smoothness and the discontinuities to be on the map of the disease in question. It assumes that the log relative risk can be written as follows.

$$Log(\theta_{ij}) = \alpha_j + v_{ij} + p_{ij}u_{ij} + (1 - p_{ij})\varphi_{ij}.$$

where v_{ij} follows a Normal distribution ($v_{ij} \sim N(0, 1/\tau_v^2)$), and p_{ij} follows a Beta distribution ($p_{ij} \sim Beta(a, a)$). While u_{ij} is a component representing unstructured heterogeneity to measure the variation in an individual area. The two mixing components are u_{ij}, $i = 1,2,..,h$, a spatial correlation, and ϕ_{ij} , $i = 1,2,..h$, a component that models discrete jumps. Special cases of this formulation arise depending on the value of P_{ij} (if it is equal to zero, we obtain the BYM model and if it is equal to one, gives a pure jump model), when

$$P_{ij} \forall i = \begin{cases} 1 & BYM \text{ model} \\ 0 & \text{pure jump model} \end{cases}$$

Most researchers noticed that the maps produced by the mixture model were very clear and visually closer compared with maps produced by the SMRs and those produced by the BYM model (Lawson and Clark, 2002; Lawson et al., 2003).

Shared Component Model

This study used the shared component model to model the spatial variation incidence rates of the two cancers in which they share overweight or obesity as a latent spatial component. In addition, we submitted the joint modelling, which Knorr-Held and Best (2001) proposed should be applied for two disease settings.

The main advantage of the shared common model is the latent component that acts as surrogate for geographical variation of the unobserved spatially structured risk factor that affects the two diseases. The Bayesian shared component model to analyse the spatial distribution of incidence rates of the two cancers jointly was used. We considered overweight as a risk factor.

For more clarification, we suppose that the log relative risk $\kappa_{i1} = \log(\theta_{i1})$; $\kappa_{i2} = \log(\theta_{i2})$ in district i for both diseases or outcomes. In other words, where the unknown parameters θ_{i1} and θ_{i2} are the log relative risk for two diseases or cancers (i.e. for first cancer and second cancer in district *i*). Then, we suppose that

$$O_{i1} \sim poisson(\mu_{i1} = E_{i1} \exp(\kappa_{i1}));\ O_{i2} \sim poisson(\mu_{i2} = E_{i2} \exp(\kappa_{i2}))$$

Here, in the previous form the Poisson mean include of a product of the relative risk ($\exp(\kappa_{i1})$ *and* $\exp(\kappa_{i2})$ and expected count (E_{i1} and E_{i2}) in district *i* for cancer 1 and 2, respectively. Where O_{i1} and O_{i2} ($i = 1,2, ..., 22$) are the disease counts or observed number of admissions by disease 1 and 2 respectively, While, E_{i1} and E_{i2} ($i = 1,2, ..., 22$) the expected number of cases for both disease 1 and 2 respectively. We modelled the log relative risk for two outcomes as below:

$$\kappa_{i1} = \log(E_{i1}) + \alpha_1 + m_{i1};\quad \kappa_{i2} = \log(E_{i2}) + \alpha_2 + m_{i2}$$

Where, the parameters α_1 and α_2 are the disease specific intercept. Here, in this model, those spatial structure may be introduced and presented to a log scale by those joint structure of m_{i1} and m_{i2}, which are as follows:

$$m_{i1} = Z_i\, \delta_1 + \varepsilon_{i1};\quad m_{i2} = Z_i\, \delta_2 + \varepsilon_{i2}$$

Z_i is the shared component common or risk factors to both diseases (cancers) in district *i*. Not that this plan to this model or tis formal is unique and different in relation to that recommended and proposed in Held et al. (2005), because they have not been using their shard components common for two diseases at same time. In other word, the first shared component common was to both diseases, while the second shared component common was only relevant to disease 1 (no.1). While, the contribution of the shared component to the overall relative risk is weighted by the scaling parameters δ_1 and δ_2 to allow a different risk gradient (on the log-scale) to be the included terms. Finally, ε_{ij} are the disease specific heterogeneous effects to capture possible variations not explained by the terms included in the model (Held et al., 2005; Downing et al., 2008).

In the Bayesian model, all unknown parameters (whether fixed or random effects) are given prior distributions. We want priors that combine both BYM

and mixture framework to link risk in space. For the joint spatial random effects, Z_i , we supposed 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 neighbouring areas to capture local dependence in space. Furthermore, a flat prior was assigned to the cancer specific intercepts, α_j . Independent normal prior distributions were used for the logarithms of the scaling parameters, log δ . According to Richardson et al. (2006), we independently assigned a conjugate hyperprior gamma (0.5, 0.0005) distribution to the precision of the shared component, τ , which is weakly informative. In the end, the disease specific heterogeneity random effects, ε_{ij}, were assigned a multivariate normal prior distribution with covariance matrix P to allow for correlations among the cancers. The inverse of this matrix known as a precision matrix, Σ^{-1} modelled to arise from a Wishart (Q,2) prior distribution, where Q is set to be a diagonal matrix with 1s (see, for example, Manda et al., 2011; Downing et al., 2008; Best and Hansell, 2009)).
All the models considered in this study were fitted to the data using full Bayesian estimation using WinBUGS version 3.2.2 software, which is a package designed to carry out a wide variety of Bayesian models. For the joint model, all fixed effects, weight, and variance parameters for convergence were monitored. According to Spiegelhalter et al. (2002a), the Brooks-Gelman-Rubin diagnostic tool, which confirmed rapid convergence by 10,000 iterations, was used in this study and inference on a chain length of 10,000 after convergence was based.

2.4 Criteria Compare the Models

In this study, we evaluate the models goodness of fit (GOF) measures to help us determine which model to be most appropriate. The use of GOF measures is common in statistical modelling to compare fitted models. Lawson (2009) showed that there are many methods that can be used as model GOF measures such as Bayes information criterion (BIC), posterior predictive, Akaike information criterion (AIC) and deviance information criterion (DIC). In this article, the last measure is used as model GOF measure because it is appropriate for use with Bayesian hierarchical models and can be evaluated easily in the WinBUGS software. The DIC was proposed by Spiegelhalter et al. (2003) and can be defined as:

$$DIC = \overline{D} + p_D = 2\overline{D} - D(\overline{\Theta}),$$

where $D(\overline{\Theta})$ is the deviance evaluated at the posterior expectation, $\overline{D}$ is the posterior expectation of the deviance and p_D is the effective number of parameters. As with all the other likelihood criteria, the DIC included penalty for the increasing complexity of the model and represents the goodness of fit. The model with the smallest DIC is estimated to best predict a replicate dataset of the same structure as that currently observed (Spiegelhalter et al., 2003).

According to (Spiegelhalter et al., 2003), the model with the lowest DIC value is estimated to be the model that would best predict a replicate dataset of the same structure as that currently observed**.** Lawson et al. (2003) pointed out that the overall goodness of fit measures **is** useful for helping model selection. They give little help in assessing how well the model fits the data. The DIC value of the multivariate modelling of the two cancers using overweight as the shared component was compared to the sum of the DIC values from the two individual BYM and mixture models. Finally, all disease maps were created with the geographical information system (GIS).

Application of BYM, Mixture and Joint Models to two Cancers Mapping

This section explains the outcomes of the applications of the joint model with existing relative risk estimation methods, corresponding to the BYM and mixture models using observed colon and oesophagus data of Libya. The data set are analysed using Win BUGS software. The findings are then compared and presented in table, maps, and DIC as measure of GOF, and the best fitted model for relative risk estimation for two cancers mapping in Libya is disclosed.

3.1 The Data Set

According to a 2011 report in Libya, the total population was 5,922,000 people. The highest number of people was in the capital of city 'Tripoli' was 1,101,000 and the minimum number was 32,000 in the district of Ghat (table 1, Fig. 1). The three models are applied to data for the two cancers in the form of the number of cases within 22 administrative districts in Libya from 2015 to 2020.

Table 1. Names of the Former 22 Districts of Libya and Their Corresponding Area Number.

Area No.	Area Name	Area Population (2011)	Area (km^2)	Area No.	Area Name	Area Population (2011)	Area (km^2)
1.	Alnikat	300,000	6,089	12.	Albatnan	169,000	84,996
2.	Zawia	302,000	2,753	13.	Nalut	101,000	67,191

3.	Aljafara	454,000	2,666	14.	Aljabal Algarbi	322,000	76,717
4.	Tripoli	1,101,000	835	15.	Wadi Shatee	81,000	97,160
5.	Almergaib	457,000	6,796	16.	Aljufra	71,000	117,410
6.	Musrata	567,000	29,172	17.	Ejdabiya	195,000	105,523
7.	Sirt	149,000	225,437	18.	Ghat	32,000	68,482
8.	Benghazi	681,000	11,372	19.	Wadi Alhiya	79,000	31,485
9.	Almarg	194,000	13,515	20.	Sabha	133,000	107,310
10.	Aljabai Alakhader	216,000	11,429	21.	Morzuk	81,000	356,308
11.	Darna	173,000	31,511	22.	Alkufra	64,000	433,611
Total	-	5,922,000	1,887,768 km^2	-	-	-	-

ID	Districts	ID Neighbours
1	Alnikat	2,13
2	Zawia	1,3,14
3	Aljafara	2,4,5,14
4	Tripoli	3,5
5	Almergaib	6,14
6	Musrata	5,7,14
7	Sirt	6,14,16,17
8	Benghazi	9,17
9	Almarg	8,10,17
10	Aljabai Alakhader	9,11,17
11	Darna	10,12,17
12	Albatnan	11,17
13	Nalut	1,14,15
14	Aljabal Algarbi	2,3,5,6,7,13,15,16
15	Wadi Shatee	13,14,16,18,19,20
16	Aljufra	7,14,15,17,20,21,22
17	Ejdabiya	7,8,9,10,11,12,16,22
18	Ghat	15,19,21
19	Wadi Alhiya	15,18,20,21
20	Sabha	15,16,19,21
21	Morzuk	16,18,19,20,22
22	Alkufra	16,17,21

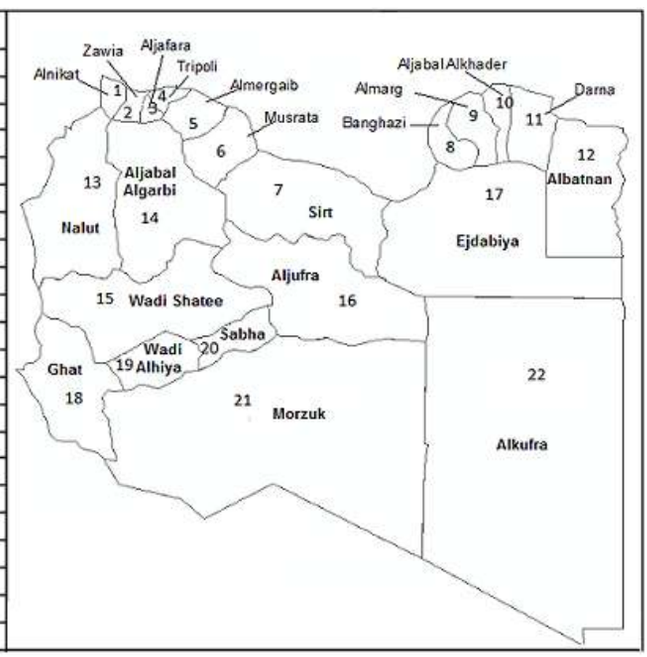

Figure 1. Study Area and Map of the 22 Name of Districts of Libya with its ID and neighbours (Geographic boundaries of districts in Libya).

Table 2 shows the range and mean of the SMRs for colon and oesophagus cancers. While table 3 presents the correlations among the SMRs. It can be seen from table 2 that cancer of the oesophagus was more common than expected in the country. The correlation among the SMRs was higher for colon and oesophagus cancer, with value equal 0.6771.

Table 2. Mean and range of the standardised morbidity ratios for two cancers

	Maximum SMR	Minimum SMR	Mean SMR
Colon	4.8134	0	0.9851
Esophagus	5.9181	0	1.2103

Table 3. Correlations between the standardised morbidity ratios for two cancers

	Colon	Esophagus
Colon	1	
Esophagus	0.6771	1

3.2 The Findings

Our analysis is related to the incidence rates of colon and oesophagus cancers from 2015 to 2020. The result reported the relative risk estimates of these cancers jointly with overweight as a risk factor (shared component). Therefore, the outcomes of relative risk estimation for three models in all districts in Libya are displayed in table 4, for colon and oesophagus cancers, respectively. This table presents numerical values for the relative risk based BYM model, mixture model, and joint mode. Our results corresponding to the joint model demonstrate the smallest range of posterior expected relative risk across districts when compared with the other two models, with maximum values of 3.775 and 4.59, and minimum values of 0.1109 and 0.06755 for colon and oesophagus cancers, respectively.

Table 4. Comparison of the Relative Risk Estimation Based on BYM model, Mixture Model and Shared Component Model for the Years 2015/ 2020

Posterior Expected Relative Risks for two Cancers Mapping						
Model	BYM		Mixture		SC	
Districts	Colon	Esophagus	Colon	Esophagus	Colon	Esophagus
Alnikat	2.086	5.372	3.495	15.29	2.107	4.786
Zawia	4.757	3.758	8.151	10.71	0.6279	1.364
Aljafara	0.6078	0.7335	1.022	1.893	0.444	0.3263
Tripoli	1.376	0.4464	2.328	1.234	0.794	0.173
Almergaib	0.4602	0.2928	0.8085	0.4711	0.1109	0.06824
Musrata	0.3484	0.7649	0.6148	2.236	0.1797	0.06755
Sirt	0.9745	0.7326	1.64	2.48	1.181	1.381
Benghazi	0.1428	0.09655	0.2502	0.3719	1.06	1.979
Almarg	0.08187	0.1445	0.0628	0.7472	0.3158	0.1996
Aljabai Alakhader	0.1138	0.1493	0.1649	0.7048	1.292	2.985
Darna	0.1818	0.1618	0.354	0.8141	0.3232	0.1501

Albatnan	0.08719	0.1799	0.0686	0.8085	4.591	4.331
Nalut	1.208	1.596	1.974	3.28	0.6183	0.4448
Aljabal Algarbi	1.35	0.6733	2.279	1.421	0.4258	0.6124
Wadi Shatee	0.9687	0.9959	1.575	1.177	0.4356	0.1696
Aljufra	1.808	0.6199	3.223	1.26	0.2801	0.1857
Ejdabiya	0.2894	0.1817	0.5574	0.7454	0.2958	0.2069
Ghat	0.3918	1.229	0.5846	1.715	0.8246	0.5832
Wadi Alhiya	1.101	3.905	1.829	12.43	0.4253	0.3605
Sabha	2.946	4.326	5.067	14.04	0.3524	0.3584
Morzuk	0.5	0.8572	0.8376	1.144	3.775	5.49
Alkufra	0.2377	0.4501	0.2689	1.322	0.4319	0.3781

It can be seen that, for both models (BYM and mixture model) the districts of Zawia and Alnikat have the highest risk of 4.757 and 8.151 for colon cancer, and of 5.261 and 15.29 for oesophagus cancer, respectively. Conversely, by using the joint model, susceptible people within the districts of Albatnan and Morzuk have the highest risk of contracting colon and oesophagus cancers, with values 4.591 and 5.49, respectively. While susceptible people within the districts of Almergaib and Musrata have the lowest risk of two cancers, when compared with people in the overall population. The corresponding values of relative risk are approximately 0.1109 and 0.06755, respectively. Furthermore, we confirm that by using the classical method the value of relative risk becomes zero when there is no observed count in a particular district, which is an acknowledged drawback of SMR. None of the other three models suffers from this drawback. More importantly, the multivariate spatial model with risk factor the we developed, is more appropriate than the other models because it has potential benefits of a joint disease mapping, such as ease of interpretation, ability to identify shared and specific patterns of risk among different diseases, improvement in the precision of the underlying disease pattern estimation, and improvement in GOF measures evaluation criteria, whereas the alternative models do not consider this level of detail.

All the results above in the table 4 might be displayed in maps which represent the high and low risk areas of two cancers occurrences. These maps give a clear picture of which district has high risk of these cancers and could be used as a tool to identify districts that need closer scrutiny or further attention in terms of government policy and financial support. Therefore, a comparison of the maps using three different methods is also made to help ascertain which method produces a smoother map.

3.3 Maps of Relative Risk of Cancers Disease in Libya from 2015 to 2020 using BYM, Mixture, SC Models

The objective of using disease mapping is to investigate the geographical distribution of the risk of a certain disease. Therefore, in this section, the results of estimation of the common models and joint model are displayed in maps in order to investigate the high and low areas of the sampled cancers. Fig. 2 and Fig. 3 show the thematic cancer risk maps for relative risk estimation based on the BYM model, the mixture model, and multivariate shared component model for two cancers in the 22 districts of Libya. Each district is assigned one of five different classes of risk which are very low risk, low risk, medium risk, high risk and very high risk, with respective intervals of [(<0,5), [0.5,1), [1,1.5), [1.5,2) and (>=2)], which were selected to cover the range of observed values with five suitable categories, based on the definition of relative risk. In addition, several colours are applied to illustrate different levels of risk for all thematic maps. The darkest hue represents the very high risk and the lightest shade representing the very low risk for different levels of relative risk for all choropleth maps.

In this study, the individual maps of colon and oesophagus cancers are presented in Fig. 2. This figure shows the overall pattern of the relative risk estimates from the BYM model and mixture model for two cancers. It can be seen from Fig. 2A and Fig. 2B based on BYM model for colon and oesophagus cancers, that the relative risk of colon cancer is higher in the western part of the country and in the centre. The focus of the highest incidence rate is in Zawia (>=2) (see Fig. 1). While, the districts with lowest risk were found in the northern and eastern parts of the country (<0.5). Furthermore, Fig. 2B presented the pattern of relative risk for oesophagus cancer. It shows that four districts have very high incidence, but the concentration of the highest incidence was found in Alnikat (1->=2). For the mixture model, Fig. 2C and Fig. 2D show the overall posterior relative risk surface for colon and oesophagus cancers, respectively. Fig. 2C for colon cancer shows that there are several districts that have high risk, which were found in the western part of the country and some districts in the centre. Likewise, Fig. 2D for oesophagus cancer shows that the cancer incidence risk distributed in total districts, but the concentration of high incidence is partly in the northwest. Comparisons between the BYM map and mixture map demonstrate that the western half of the country and some districts in the centre were more in high risk than the other districts for each cancer. Conversely, these maps show that the southern and eastern pasts of the country have low relative risk for both cancers.

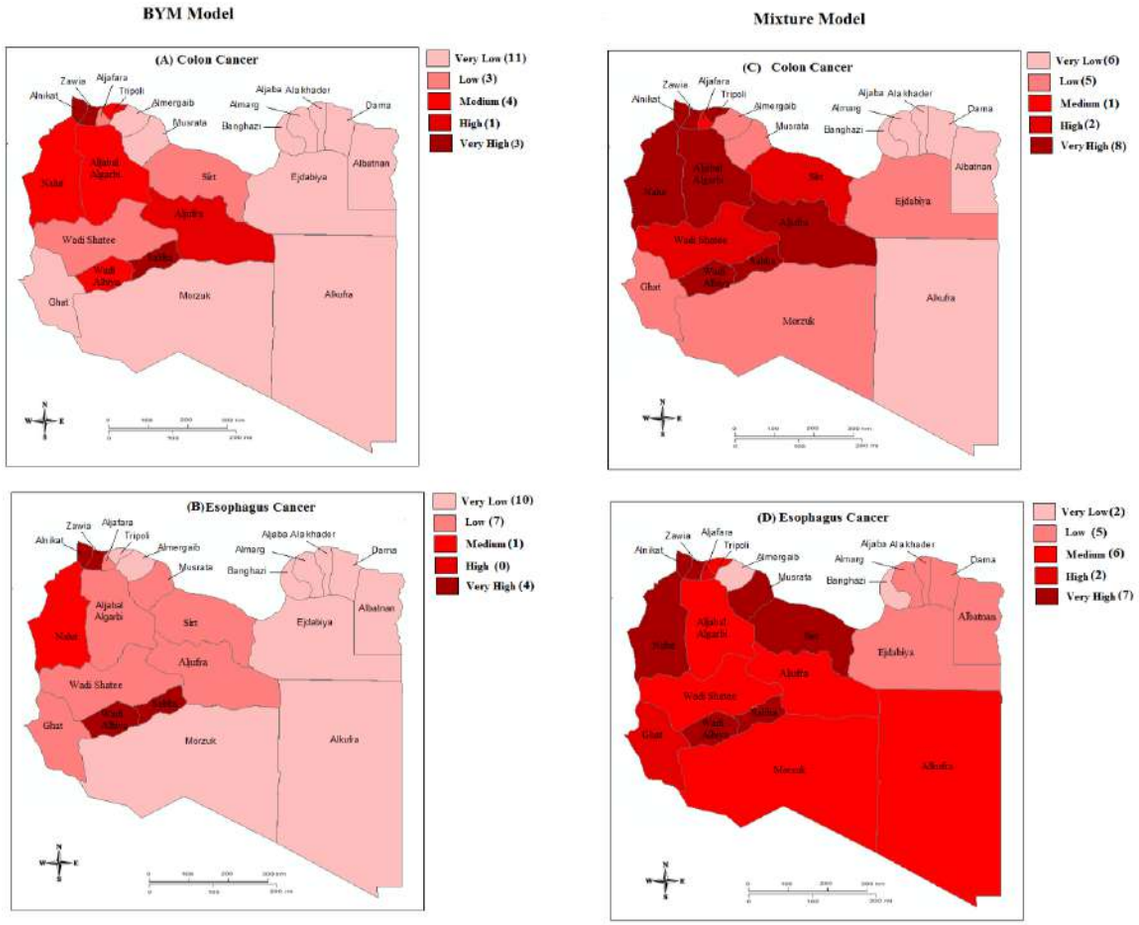

Figure 2. Maps of the posterior Estimated Relative Risk in the BYM model and Mixture Model for Colon and Esophagus Cancers in Libya, 2015/2020.

In addition, our results demonstrated the relative risk estimates of colon and oesophagus cancers with overweight as a risk factor. Furthermore, we display the multivariate modelling of two cancers in women and men, individually. The estimates of the effects of the risk factor or shared component were also mapped in Fig. 3. Fig. 3A shows the overall posterior relative risk surface of joint analysis for colon and oesophagus cancers with obesity from 2015 to 2020. This map contains dark pink and crimson, which denotes the incidence rate is 0.5-1.5. According to the derived map in this figure, overweight had more effect in the north-western districts. These districts Almergaib, Musrata and Aljabal Algarbi, and only one district in the east, which is Ejdabiya, one located in the centre part in the country, which is Aljufra, and in Ghat, Sabha, Wadi Shatee and Wadi Alhiya in the south-west. Fig. 3B and Fig. 3C explain the posterior relative risk surface of multivariate analysis for women and men, respectively. These maps have the same pattern as the general map, except the districts of Zawia and Nalut

for women and men, respectively. These districts show medium risk for colon and oesophagus incidence. However, for women and men the distribution of incidence rate is a little different only in two districts as shown in Fig. 3B and Fig. 3C. The controlling feature of the general multivariate map is an increasing trend from the south-west to the west.

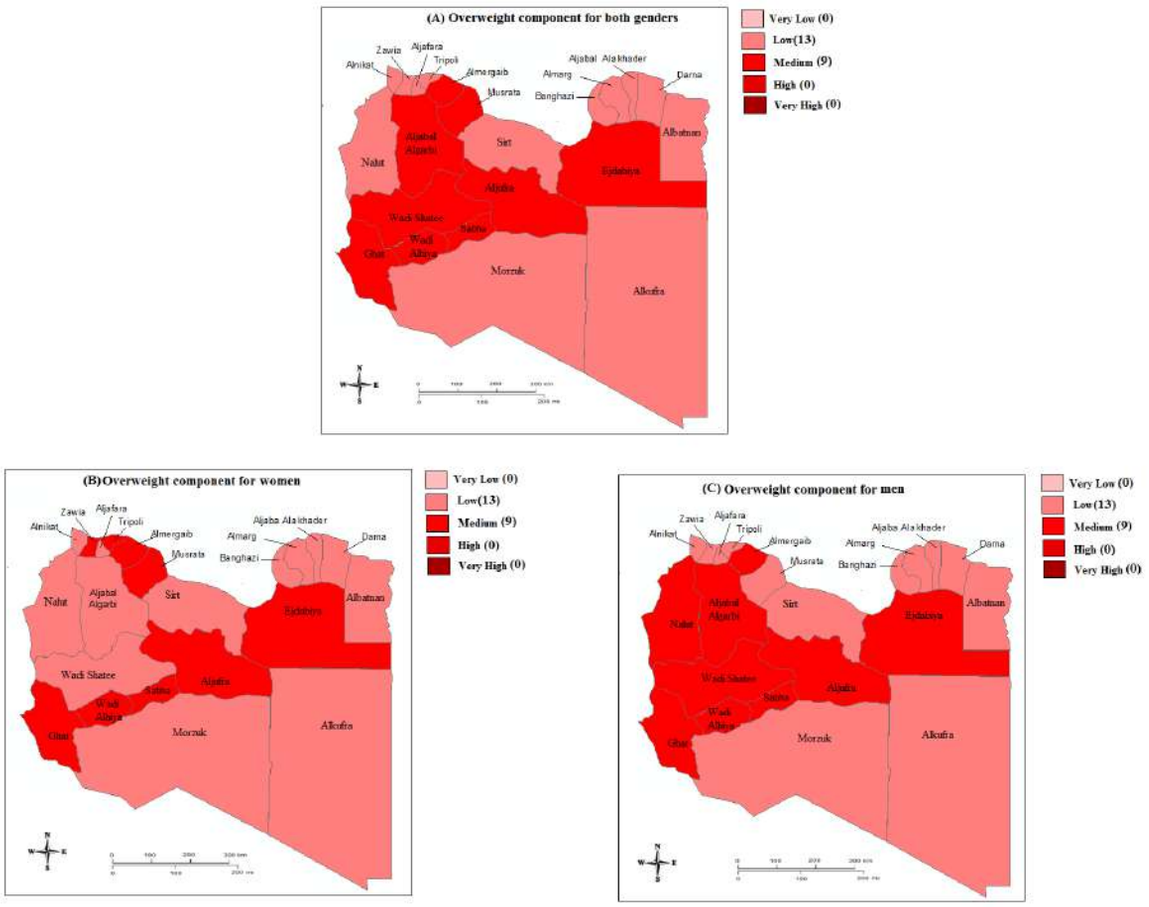

Figure 3. Maps of the posterior Estimated Relative Risk for Colon and Esophagus Cancers in Libya, 2015/2020, using joint model.

3.4 Deviance Information Criterion Model Selection

In this study, the DIC was used to compare the variations of the multiple models with BYM model and mixture model. Table 4 shows the DIC values for relative risk estimation for the two cancers incidence for all districts in Libya from 2015 to 2020. From the DIC values in table 5, we conclude that the model with SC model fits best because it gives the smallest DIC compared to the other models. From the DIC values in table 5, a comparison of the DIC showed that the mixture model is much better than the BYM model, which has the largest DIC value among all models. We conclude that the model with multiple diseases fits best because it gives the smallest DIC compared to the other models. This

displayed a great improvement in DIC values for multiple models. i.e., this shows a conclusion that the joint model which assumes that the data for colon and oesophagus cancers is the best model to be used in the analysis specifically for estimation of relative risk. This clearly shows that DIC joint modelling of two cancers has a feature over modelling them individually. As a result, it is suggested that this is the most robust and appropriate model to be used.

Table 5. Goodness of fit (Deviance Information Criterion DIC) for BYM model, Mixture and joint models to estimate relative risk of two cancers.

	Model		
Cancer/ DIC	BYM	Mixture	SC
Colon	122.398	118.666	-
Esophagus	60.699	60.846	-
Total	183.097	179.512	**174.154**

Conclusion

In this study, the major goal was applying the multivariate shared component model to analyse the joint spatial distributions of colon and oesophagus cancers incidence rates from 2015 to 2020. The features and advantages of spatial analysis of disease rates have been provided, as well as the aim of joint modelling of different diseases and its benefits, the multivariate shared component model structure, the data sources, and assumptions and formulation. Two types of cancers are included in the joint model as response variables in relation to an overweight, as a risk factor, which caused these cancers.

The maps presented the geographical differences in two cancers incidence rates, as well as the high and low risk districts in Libya. In addition, the general multivariate map displayed that the western half of the country was at a higher risk than the southern and eastern half. This pattern remained for male and female maps, but for Zawia and Nalut, the relative risk estimate was different. In our analysis, the joint model offers ease of interpretation. A joint model of the two cancers achieves a considerable improvement in terms of DIC over the most common individual modelling of diseases, BYM and mixture. This kind of application or analysis might be helpful for governmental authorities to appreciate the health care system performance and set appropriate policies. In this research, the geographical pattern of relative risk using a multivariate shared component model indicates that there are several risk factors, such as tobacco use and alcohol abuse component, which are important in the country and more awareness is needed in the allocation and delivery of public health policies. In other words, our application used an obesity as a shared component or risk

factor, but we can confirm that the other risk factors common for these cancers such as smoking, low physical activity, diet low in fruit and vegetables, and others, must receive more care in the high-risk districts. Finally, we are working on a multivariate model for disease mapping of these two cancers in addition to five other cancers and four risk factors in Libya using a shared component model to show the spatial pattern of the diseases and to demonstrate the advantage, feasibility, and utility of the shared component model in multivariate spatial analyses. We also compared it with several common models used in disease mapping.

Acknowledgments

The author would like to thank the staff members of the African Oncology Institute (AOI), Sabratha at the Libyan Cancer Centre for their continued support for completion of this study.

References

- Aburwais A, Elgonoti S, Rafael A, Boder J, et al. (2012): Demographic Features of Common Cancers in the National Cancer Institute, Misurata, Biomedical Journal of Scientific & Technical Research, Biomedical Research Network+, LLC, vol. 38(1), pages 29936-29943, August.
- Alramah, M. A., Samat, N. A., & Mohamed, Z. (2019): Mapping lung cancer disease in Libya using Standardized Morbidity Ratio, BYM model and mixture model, 2006 to 2011: Bayesian Epidemiological Study. Sains Malaysiana, **48(1)**, 217-225.
- Assucão RM, (2004): Castro MS. Multiple cancer sites incidence rates estimation using a multivariate Bayesian model. Int J Epidemiol., 33(3):508-16.
- Besag J, York J, Mollie´ A. (1991): Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math., 43(1):1-20Kk
- Best N, Hansell AL (2009): Geographic variations in risk: Adjusting for unmeasured confounders through joint modeling of multiple diseases. Epidemiology., 20, 400-10.
- Dabney AR, Wakefield JC. (2005): Issues in the mapping of two diseases. Stat Methods Med Res., 14(1):83-112.
- Downing A, Forman D, Gilthorpe MS, Edwards KL, Manda SO. (2008): Joint disease mapping using six cancers in the Yorkshire region of England. Int J Health Geogr., 28;7:41.
- Dreassi E. (2010): 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.
- El Mistiri M, Pirani M, El Sahli N, El Mangoush M, Attia A, Shembesh R, Habel S, El Homry F, Hamad S, Federico M. (2010): Cancer profile in Eastern Libya: incidence and mortality in the year 2004. Ann Oncol., **21**: 1924-1926.

- Held L, Nata´rio I, Fenton SE, Rue H, Becker N. (2005): Towards joint disease mapping. Stat Methods Med Res., 14(1):61-82.
- Knorr-Held L, Best NG. (2001): A shared component model for detecting joint and selective clustering of two diseases. J R Stat Soc Ser A Stat Soc.,164(1):73-85.
- Langford IH, Leyland AH, Rasbash J, Goldstein H. (1999): Multilevel modelling of the geographical distributions of diseases. J R Stat Soc Ser C Appl Stat., 48(2):253-68.
- Lawson A, Biggeri A, Boehning D, et al. (2000): Disease mapping models: an empirical evaluation. Disease Mapping Collaborative Group. Stat Med., 19(17-18):2217-41.
- Lawson, A.B. (2009). Bayesian disease mapping: hierarchical modelling in spatial
- Lawson, A.B. and Clark, A. (2002): Spatial mixture relative risk models applied to disease mapping, Statistics in Medicine, 21, pp, 359-370.
- Lawson, A.B., Browne, W.J. and Vidal Rodeiro, C.L. (2003): Disease mapping with WinBUGS and MLwiN. England: John Wiley & sons.
- Leyland AH, Langford IH, Rasbash J, Goldstein H. (2000): Multivariate spatial models for event data. Stat Med., 15-30; 19(17e18):2469-78.
- Libyan National Statistics Figures: Annual Statistical Report. Tripoli, 2011. (In Arabic).
- Manda SM, Feltbower RG, Gilthorpe MS. (2011): Review and empirical comparison of joint mapping of multiple diseases. S Afr J Epidemiol Infect., 27(4):169-82.
- Richardson S, Abellan JJ, Best N. (2006): Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire (UK). Stat Methods Med Res., 15(4):385-407.
- Sabratha Cancer Registry: First annual report, 2006. Edited by African Oncology Institute, Sabratha, Libya, 2008.
- Spiegelhalter DJ, Thomas A, Best NG, et al. (2002a): WinBUGS Bayesian inference using Gibbs sampling manual. Version 1.4. Imperial College London and MRC Biostatistics Unit, Cambridge. Available at: http:/www. mrc-bsu.cam.ac.uk/bugs
- Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003): WinBUGS user manual version 1.4. Cambridge UK: MRC Biostatistics Unit.
- Tzala E, Best N. (2000): Bayesian latent variable modelling of multivariate spatio-temporal variation in cancer mortality. Stat Med., 15:30;19(17-18):2217-41.