EMPIRICAL AND FULLY BAYESIAN METHODS FOR IDENTIFYING THE SPATIAL DISTRIBUTION OF THE ADOLESCENT PREGNANCY RATES IN A BRAZILIAN STATE

Daiane Leite da ROZA[1]
Edson Zangiacomi MARTINEZ[1]

▪ *ABSTRACT: Spatial statistical methods are often employed to describe the spatial distribution of the occurrence of an event of interest in a geographical space. In this paper we compare the global and local empirical Bayesian methods and the fully Bayesian method for identifying the spatial distribution of the adolescent pregnancy rates in Minas Gerais, the largest State in the Southeastern of Brazil. The fully Bayesian method consider a conditional autoregressive (CAR) distribution to the spatial components of the model. The methods used in this study suggest that the percentages of live births to adolescent mother (LBAM) are not randomly distributed across the State of Minas Gerais. The percentages of LBAM tend to be higher in the North than in the South of the State. This suggest an important association between the occurrence of teenage pregnancy and socio-economic indicators, given that the North is the poorest region of the State while the South region concentrates the municipalities with the highest rates of development.*

▪ *KEYWORDS: Spatial statistical methods; Bayesian methods; pregnancy in adolescence; ecological study.*

1 Introduction

In the recent literature, spatial statistical methods has been widely used in studies in the health sciences to describe the spatial pattern in the distribution of disease occurrence or other events of interest (WALLER & GOTWAY, 2004; LAWSON, 2013). These methods basically consider two type of data (ELLIOTT & WARTENBERG, 2004). Point data are geographically indexed data, representing

[1]University of São Paulo – USP, Ribeirão Preto Medical School, Department of Social Medicine, CEP: 14049-900, Ribeirão Preto, SP, Brazil. E-mail: *daianeroza@usp.br*; *edson@fmrp.usp.br*

the spatial location of the occurrence of the event of interest. These points can be represented, for example, as pairs of latitude and longitude coordinates. Area data are defined as data collected with respect to a given spatial region such as a community, municipality, census sector or the coverage area of a health district. The point process statistics analyse the geometrical structure of patterns formed by locations of occurrence of events distributed in a space (ILLIAN et al., 2008). These methods include the description of the different types of correlations in the patterns. Many statistical models for continuous or discrete valued area data include a random effect which is conditioned by the neighbourhood structure (BAILEY, 2001; CARVALHO & SOUZA-SANTOS, 2005). The fit of these models is usually based on Bayesian estimation including classes of conditional autoregressive (CAR) prior distributions.

In a general manner, ecological studies are defined as those in which the unit of the analysis are aggregated or grouped observations, rather than individuals (LAST, 2001). The objective of a ecological study is thus to provide informations about the associations between potential risk factors and the observed incidence of a disease or a event of interest as measured on groups. Considering that these groups are typically defined by geographical areas, spatial statistical methods are especially useful for these purposes. As a special case, a number of ecological studies have been published on the environmental influences that may determine the adolescent pregnancy. As examples we can cite the studies of Hau et al. (2009), Nogueira et al. (2009), Martinez et al. (2011) and Martins et al. (2014). In all of these studies, it is shown that the adolescent pregnancy rates are higher in regions with low indicators of education and socio-economic development.

Many spatial modelling techniques are developed within a Bayesian approach due its flexibility in structuring complicated models and ease of analysis (AGUERO-VALVERDE & JOVANIS, 2006). The estimation of the parameters of models based on the CAR distribution and their generalizations can be a difficult task when frequentist inference methods are used, due to the complexity of the likelihood function (MARTINEZ & ACHCAR, 2014). In this case, Bayesian methods provide a convenient alternative to deal with this model structure. In epidemiological studies that attempt to account for associations between the rate of any event of interest and the geographical space in which the event takes place, the results of the Bayesian analysis are commonly presented in "smoothed maps"(CASTRO et al., 2004). These maps are constructed from smoothed rates, which are obtained by combining the knowledge on the rates of each geographical area and the rates in the neighbouring areas. In this context, empirical and fully Bayesian methods use information from other regions that comprise the study area to reduce the effects of random fluctuations that are not associated with the risk of the event of interest (ASSUNÇÃO et al., 1998). Thus, the resultant maps are smoother and more informative.

In the present study, we compare the empirical and fully Bayesian methods for identifying the spatial distribution of the adolescent pregnancy rates in Minas Gerais, the largest state in the Southeastern of Brazil. Minas Gerais is the second most important Brazilian State in terms of economic development and population. The State has an area of 586.522 km^2 divided into 853 municipalities, and a population of over 20 millions of inhabitants (Brazilian Institute of Geography and Statistics estimate for 2014). The State of Minas Gerais is characterized by deep social differences existing between the population in the North and South areas. The North of the State is a region with high incidence of poverty, while the more developed municipalities are located at the South region. Data on adolescent pregnancy were obtained from the Live Birth Information System (SINASC, acronym for "*Sistema de Informações sobre Nascidos Vivos*"), considering the year of 2010 as the period of study.

The rest of the article is structured as follows. In Section 2 the fully Bayesian model using a CAR spatial structure is presented. Section 2 describes the global and local empirical Bayesian methods. The results obtained from these approaches are presented in Section 3. Section 4 contains a discussion of the results and the final comments.

2 Fully Bayesian (FB) model

Let us consider that the State of Minas Gerais is partitioned into M municipalities indexed by i, $i = 1, ..., M$. The fully Bayesian (FB) model considers that for the municipality i, Y_i is the counting of live births to adolescent mother (LBAM) recorded in the year of 2010, N_i is the total number of live births in this year, and θ_i is the percentage of LBAM. For this statistical model, Y_i is a random variable following a binomial distribution with success probability θ_i in N_i independent Bernoulli trials. This is formally written by

$$Y_i|N_i, \theta_i \sim Binomial(N_i, \theta_i), \quad \text{for} \quad i = 1, ..., M,$$

where $M = 853$ is the number of municipalities of the State of Minas Gerais. We consider that N_i is known and θ_i is a parameter to be estimated by the model. In many applications of spatial models, it is very common the use of the Poisson distribution for modeling count data. A model based on the Poisson distribution considers a large number of events N_i and small probabilities θ_i. However, in the present study, it is expected that the percentages θ_i do not assume relatively low values, favoring the use of the Binomial distribution. Assuming a logit link function between the percentage θ_i of LBAM, we have the following:

$$\theta_i = \frac{\exp(\alpha + \gamma_i)}{1 + \exp(\alpha + \gamma_i)},$$

where α is an unknown parameter (fixed effects) and γ_i are random effects associated with the i-th municipality. In the FB analysis, it is considered that γ_i assumes prior spatial distribution allowing higher correlations between close areas in the geographical space. The spatial distribution adopted here follows a conditional autoregressive (CAR) model (Besag & Kooperberg, 1995), as written below:

$$\gamma_i | \{\gamma_i, i \neq j, j \in A^*(i)\}, \sigma_\gamma^2 \sim N\left(\overline{\eta}_i, \frac{\sigma_\gamma^2}{n_i}\right),$$

where $A^*(i)$ means a set of municipalities neighbour to the municipality i, $\overline{\eta}_i$ is the mean of the random effects γ_i associated with municipalities neighbour to the municipality i, n_i corresponds to the number of municipalities neighbour to the municipality i, σ_γ^2 is a variance parameter of the CAR distribution, and $N(a, b)$ denotes generically a normal distribution with mean a and variance b. The criterion used for neighbourhood definition was that of adjacency in which municipalities having boundaries with each other were considered neighbours. Furthermore, this formulation allows the inclusion of covariates in the model, but this goes beyond the scope of the present study.

For the fit of the model, the GeoBUGS module (Thomas et al., 2004) of the software WinBUGS (LUNN et al., 2000) was used. We assumed that fixed effect α follows an improper flat type prior distribution, defined for the whole set of real numbers. Also, we assumed that the parameter σ_γ^2 follows an inverse gamma prior distribution, expressed as $\sigma_\gamma^{-2} \sim Gamma(0,5; 0,0005)$, according to recommendation by Thomas et al. (2004). We will consider that these prior distributions are independent from each other. In this model, the posterior conditional distributions have some complexity in their forms, which requires the use of computationally intensive methods for simulation of the samples of these distributions. Therefore, we used the Markov Chain Monte Carlo algorithm (MCMC) to obtain samples of these distributions. One hundred thousand samples were generated from each parameter of interest after disregarding the first $10,000$ samples in order to avoid any effect of the initial values (i.e. burn-in samples). To avoid correlations between successively generated samples, it was stored each 100th simulated sample. To compare whether models with random effects for description of the spatial structure were more adequate than those without such effects, the deviance information criterion (DIC) (SPIEGELHALTER et al., 2002) was used. This criterion determines that models with low DIC values are to be preferred over those with higher values. The WinBUGS code for fitting the FB model to the data is provided in the Appendix.

3 Global and local empirical Bayesian methods

The estimates obtained by using both global and local empirical Bayesian methods were based on Marshall (1991), who denotes by $\hat{\theta}_i$ the empirical Bayesian estimate of the LBAM percentage θ_i for the municipality i. As $p_i = Y_i/N_i$ is the crude percentage of LBAM for municipality i, we can write the following conditional mean and variance,

$$E(p_i|\theta_i) = \theta_i \quad \text{and} \quad Var(p_i|\theta_i) = \frac{\theta_i}{n_i},$$

respectively. Let us suppose that θ_i has a prior density with mean $E_\theta(\theta_i) = m_i$ and variance $Var_\theta(\theta_i) = \tau_i^2$. Therefore,

$$E_p(p_i) = E_\theta\left[E\left(p_i|\theta_i\right)\right] = m_i$$

and

$$Var_p(p_i) = Var_\theta\left[E\left(p_i|\theta_i\right)\right] + E_\theta\left[Var\left(p_i|\theta_i\right)\right] = \tau_i^2 + \frac{m_i}{n_i}.$$

Marshall (1991) assumed that $\tau_i^2 = \tau^2$ and $m_i = m$ for all i, or say, the same prior mean and variance for all municipalities. The best linear Bayes estimator of θ_i, in terms of summed squared error loss ($SSEL$), is

$$\hat{\theta}_i = m + c_i(p_i - m), \quad \text{for} \quad i = 1, ..., M,$$

where c_i is called the shrinkage factor (Waller and Gotway, 2004) given by

$$c_i = \frac{Var_\theta\left(\theta_i\right)}{Var_p(p_i)} = \frac{\tau_i^2}{\tau_i^2 + \frac{m_i}{n_i}}$$

and

$$SSEL = \sum_{i=1}^{M}\left(\widehat{\theta}_i - \theta_i\right)^2.$$

The value of c_i is between 0 and 1, representing the weight given to the difference between p_i and m. Calculation of c_i considers the number of live births of the respective municipality. When this number is relatively high, the value of c_i is near to 1 and thus the estimate $\hat{\theta}_i$ becomes near to the crude percentage of LBAM (Silva et al., 2011). The value of c_i decreases when the number of live birth in the municipality i is small, thus attributing less weight to the difference between p_i and m and making the estimate $\hat{\theta}_i$ nearer to the global percentage for the State. The expression for $\hat{\theta}_i$ can also be written as follows:

$$\hat{\theta}_i = c_i p_i + (1 - c_i)m, \quad \text{for} \quad i = 1, ..., M,$$

where estimate $\hat{\theta}_i$ is understood as the weighted mean between crude percentage of LBAM of the municipality $_i$ and the percentage of the whole State of Minas Gerais (weighted by c_i).

In order to obtain the Bayesian estimates, we need values for m and τ^2. Considering an empirical Bayes estimation, these values are obtained from the data. Marshall (1991) proposes the method of moments to estimate values for m, τ^2 and c_i. Thus, the Bayesian estimates based on the whole State's global percentage as value of m, are denominated global empirical Bayesian (GEB) estimates. On the other hand, the local empirical Bayesian (LEB) estimator considers only those neighbour municipalities that share the same boundary within the area in which one wants to estimate the LBAM percent, converging into a local rather than a global mean value. The GEB and LEB estimators were obtained by using the software TerraView 4.2.2. of the National Institute of Space Research (INPE). For visual comparisons between these approaches, we generated graphs describing the sequential behaviour of the estimates generated according to the ascending order of crude LBAM percentages of the municipalities (SILVA et al., 2011).

4 Results

Posterior summaries obtained using the FB model are presented in Table 1. In the application of the MCMC method, convergence was obtained for all simulated chains. Thus, the generated sample of observations is valid for inference.

In addition, the spatial autocorrelation of the residuals of the Bayesian model was not detected. The standardized Moran's I statistic is -0.007, or essentially zero, indicating non-existence of spatial autocorrelation (LAWSON & KLEINMAN, 2005, Section 4.4). The 95% credible intervals for the parameters are obtained from the $2.5th$ and $97.5th$ percentiles of the sampled distributions. Table 1 shows the percentages θ_i of LBAM estimated for the first ten municipalities and for the last four municipalities, considering that the municipalities were added to the data set in an arbitrary order.

Table 2 displays the twenty municipalities of the State of Minas Gerais with the highest crude percentages of LBAM and the twenty municipalities with the lowest crude percentages of LBAM. The higher percentage of LBAM is founded in Santo Hipólito, located at a mesorregion called Central Mineira, which is in the central region of the State. Santo Hipólito is a very small municipality, with around 3,500 inhabitants (Brazilian Institute of Geography and Statistics, IBGE, census 2010). There were no records of LBAM in Antônio Prado de Minas, Catas Altas da Noruega, Córrego Danta and Serra da Saudade in the year of 2010. All of these four municipalities had populations of less than 3,500 inhabitants. The municipality of Serra da Saudade had only 822 inhabitants (IBGE, census 2010). Table 2 shows that a large part of the municipalities with the highest crude percentages of LBAM are in the North region of the State, while the municipalities with the lowest crude percentages are near the capital of the State (Belo Horizonte) or they are located at the South of Minas Gerais.

Figure 1 compares the spatial distribution of the crude and smoothed percentages of LBAM, obtained from the FB model. Both maps describe a spatial pattern such that the higher percentages of LBAM tend to be located in northern part of the State. However, this effect is more evident when we visualize the map of the smoothed percentages (Panel B of the Figure 1).

Table 2 also displays a brief comparison between the crude percentages of LBAM and the percentages estimated by the FB, GEB and LEB methods. We can note that, due the effect of the smoothing process, the percentages of LBAM obtained from the three statistical models are lower than the crude percentages when considered the municipalities with the highest crude percentages. On the other hand, the percentages of LBAM obtained from the statistical models are higher than the crude percentages when considered the municipalities with the lowest crude percentages.

Table 1 - Posterior summaries for the parameters of the fully Bayesian (FB) model applied to the adolescent pregnancy data

Parameter	mean	SD[1]	median	$2.5th$ percentile	$97.5th$ percentile
α	-1.45	0.0084	-1.45	-1.467	-1.434
γ_1	-0.1544	0.2113	-0.1508	-0.5780	0.2521
γ_2	0.0436	0.1343	0.0448	-0.2224	0.3044
γ_3	-0.1200	0.1375	-0.1195	-0.3914	0.1465
γ_4	0.1599	0.1924	0.1601	-0.2190	0.5354
γ_5	-0.0172	0.1349	-0.0167	-0.2820	0.2464
γ_6	0.2038	0.1239	0.2042	-0.0403	0.4466
γ_7	0.0219	0.3206	0.0251	-0.6131	0.6428
γ_8	-0.0960	0.2051	-0.0950	-0.5010	0.3032
γ_9	0.1650	0.1185	0.1664	-0.0698	0.3947
γ_{10}	0.0571	0.1781	0.0576	-0.2931	0.4040
...					
γ_{850}	0.1843	0.1941	0.1862	-0.2006	0.5584
γ_{851}	0.0525	0.1012	0.0532	-0.1482	0.2475
γ_{852}	-0.0014	0.2334	0.0013	-0.4656	0.4518
γ_{853}	-0.2146	0.2570	-0.2119	-0.7219	0.2840
σ^2_γ	0.2426	0.0231	0.2415	0.2003	0.2908
θ_1	0.1694	0.0295	0.1678	0.1162	0.2319
θ_2	0.1976	0.0213	0.1970	0.1581	0.2413
θ_3	0.1731	0.0197	0.1723	0.1368	0.2135
θ_4	0.2176	0.0327	0.2158	0.1585	0.2862
θ_5	0.1882	0.0206	0.1874	0.1502	0.2309
θ_6	0.2241	0.0215	0.2234	0.1837	0.2683
θ_7	0.1982	0.0504	0.1938	0.1125	0.3085
θ_8	0.1776	0.0299	0.1758	0.1244	0.2411
θ_9	0.2173	0.0201	0.2169	0.1794	0.2582
θ_{10}	0.2004	0.0285	0.1990	0.1489	0.2601
...					
θ_{850}	0.2217	0.0333	0.2203	0.1610	0.2907
θ_{851}	0.1987	0.0161	0.1983	0.1683	0.2310
θ_{852}	0.1923	0.0360	0.1902	0.1282	0.2692
θ_{853}	0.1621	0.0347	0.1595	0.1022	0.2377

[1]SD: standard deviation.

The maps of Figure 2 compare the spatial distribution of the smoothed percentages of LBAM considering the LEB and the GEB methods. The GEB method produces smoothed percentages that account for the mean across the whole State, while the LEB method accounts for the mean among spatial neighbours. Thus, the map that describes the smoothed percentages of LBAM considering the LEB method (Figure 2, Panel A) is closer to the map that describes the crude

Table 2 - The twenty municipalities with the highest crude percentages of LBAM and the twenty municipalities with the lowest crude percentages of LBAM. State of Minas Gerais, Brazil, 2010

	Municipality	Mesorregion	Percentages of LBAM			
			Crude	FBM[1]	GEBM[1]	LEBM[1]
Highest percentages	Santo Hipólito	Central Mineira	46.43	27.40	23.05	24.41
	Veríssimo	Triângulo Mineiro/Alto Paranaíba	44.19	26.97	13.57	15.76
	Ponto Chique	Norte de Minas	41.38	27.17	15.93	20.25
	Santa Fé de Minas	Norte de Minas	41.30	26.37	24.28	29.52
	Comendador Gomes	Triângulo Mineiro/Alto Paranaíba	40.74	23.65	21.71	24.14
	Mathias Lobato	Vale do Rio Doce	40.38	22.55	17.73	16.74
	Chiador	Zona da Mata	38.46	27.42	22.68	24.07
	Ipiaçu	Triângulo Mineiro/Alto Paranaíba	38.46	22.89	21.11	20.11
	Estrela Dalva	Zona da Mata	37.84	24.51	19.34	21.41
	Quartel Geral	Central Mineira	37.14	23.18	21.90	22.72
	Jampruca	Vale do Rio Doce	36.76	26.81	24.66	23.86
	Ibiaí	Norte de Minas	36.73	30.30	28.44	32.07
	Monte Formoso	Jequitinhonha	35.59	29.78	23.57	25.72
	Verdelândia	Norte de Minas	35.44	29.89	16.99	21.46
	Umburatiba	Vale do Mucuri	35.00	25.95	18.11	19.53
	Nacip Raydan	Vale do Rio Doce	34.48	23.21	20.64	23.24
	Carvalhos	Sul/Sudoeste de Minas	34.29	23.74	18.53	16.90
	Juvenília	Norte de Minas	34.29	23.43	17.51	14.38
	Cristália	Norte de Minas	34.02	26.52	16.29	14.37
	Matias Cardoso	Norte de Minas	33.55	29.74	26.72	28.62
Lowest percentages	Vermelho Novo	Zona da Mata	8.06	14.67	17.15	16.56
	Cachoeira da Prata	Metropolitana de Belo Horizonte	7.89	14.58	14.44	17.69
	Jesuânia	Sul/Sudoeste de Minas	7.50	16.03	14.25	19.76
	Santa Bárbara do Tugúrio	Campo das Vertentes	7.41	14.42	13.62	14.03
	Liberdade	Sul/Sudoeste de Minas	7.32	17.14	16.62	17.36
	Japaraíba	Central Mineira	6.78	14.30	13.20	14.02
	Confins	Metropolitana de Belo Horizonte	6.58	15.07	14.77	15.24
	Frei Lagonegro	Vale do Rio Doce	5.88	15.95	13.24	16.92
	Capela Nova	Campo das Vertentes	5.80	17.95	21.29	21.60
	Rio Manso	Metropolitana de Belo Horizonte	5.13	13.77	13.65	12.84
	Arapuá	Triângulo Mineiro/Alto Paranaíba	5.00	15.86	14.92	15.91
	Goianá	Zona da Mata	5.00	15.76	13.56	14.68
	Santa Rita de Jacutinga	Zona da Mata	4.55	17.60	16.31	19.16
	São Pedro dos Ferros	Zona da Mata	4.35	17.78	24.87	27.03
	São Sebastião do Rio Verde	Sul/Sudoeste de Minas	4.35	16.80	19.15	17.24
	Bonfim	Metropolitana de Belo Horizonte	3.64	13.16	12.28	11.64
	Antônio Prado de Minas	Zona da Mata	0	16.36	17.92	18.67
	Catas Altas da Noruega	Metropolitana de Belo Horizonte	0	21.29	15.76	22.46
	Córrego Danta	Oeste de Minas	0	16.65	14.84	17.49
	Serra da Saudade	Central Mineira	0	13.08	12.26	10.73

[1]FBM: Fully Bayesian model, GEBM: global empirical Bayesian method, LEBM: local empirical Bayesian method.

percentages (Figure 1, Panel A). The plots in Figure 3 compare the smoothed percentages obtained from the three Bayesian methods with the crude percentages of LBAM. These plots suggest an excessive smoothing in municipalities where the percentages of LBAM are relatively high. This effect is common in applications of spatial statistical methods and it is discussed by Castro et al. (2004).

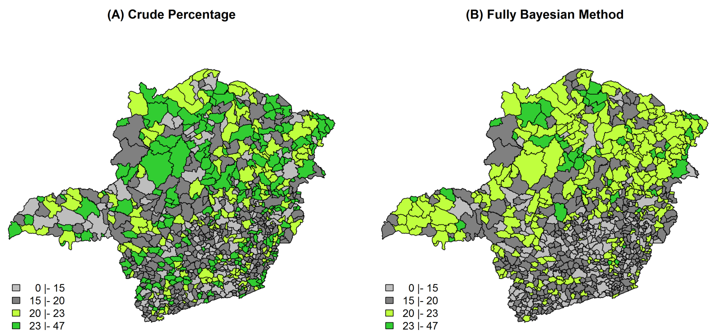

Figure 1 - Spatial distribution of the (A) crude and (B) smoothed percentages of LBAM, considering the fully Bayesian (FB) model. State of Minas Gerais, Brazil, 2010.

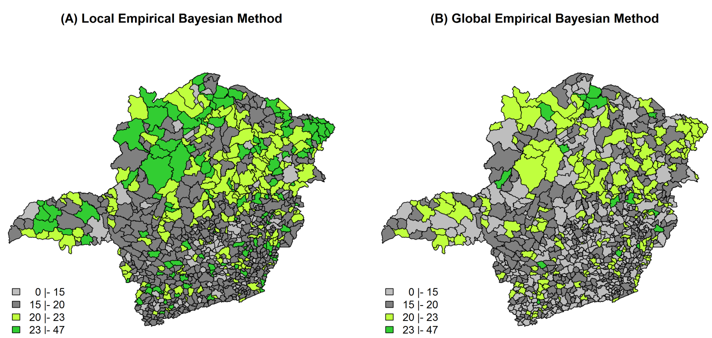

Figure 2 - Spatial distribution of the smoothed percentages of LBAM considering the (A) local empirical Bayesian (LEB) method and the (B) global empirical Bayesian (GEB) method. State of Minas Gerais, Brazil, 2010.

Figure 4 shows histograms of the crude percentages of LBAM and the smoothed percentages estimated from the FB, LEB and GEB methods. The Panel A of the Figure 4 demonstrates a skew distribution of the crude percentages of LBAM, ranged from 0 to 46.43%. In addition, the Figure 4 illustrates that the dispersion of the percentages of LBAM is smaller when estimated using the Bayesian methods than they are directly obtained from the data.

Figure 5 describes the percentages of LBAM estimated by using different methods, always in ascending order according to crude percentages. This figure is similar to one from Silva et al. (2011). It was observed that estimates obtained from the empirical Bayesian method, both local and global, had values near to results found by the FB method. The GEB method has smoother estimates and uses the global mean percentage as reference, thus presenting a “funnel” at the centre

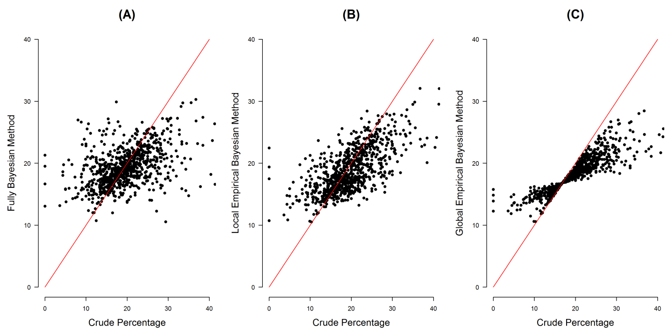

Figure 3 - Plots comparing the smoothed percentages obtained from the (A) fully Bayesian (FB), (B) local empirical Bayesian (LEB) and (C) global empirical Bayesian (GEB) methods with the crude percentages of LBAM.

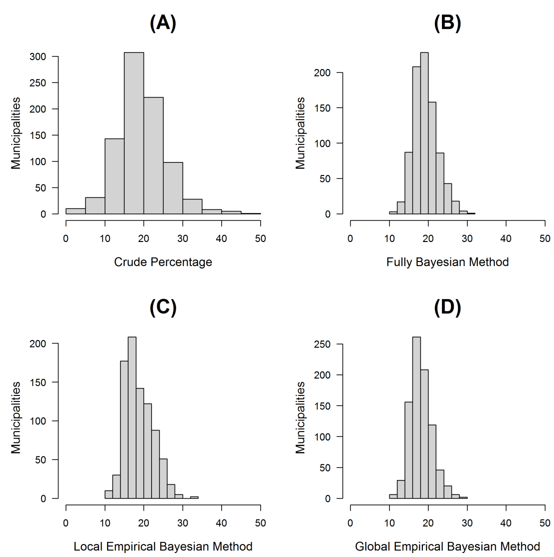

Figure 4 - Histograms of the (A) crude percentages of LBAM and the smoothed percentages estimated from the (B) fully Bayesian (FB), (C) local empirical Bayesian (LEB) and (D) global empirical Bayesian (GEB) methods.

of the estimate graph (Figure 5, Painel B). For estimates obtained by using the LEB method, the variability is higher as the values remained high in some municipalities (Painel C). We can observe that values estimated by using the LEB method is more similar to the results obtained from the FB method (Painel D), since they have a greater heterogeneity than the GEB approach. Also, it was observed that graphs no longer presented percentages equal to zero for the municipalities after smoothing, since the estimated percentages had incorporated the implicit pattern of the region in which they are inserted.

P

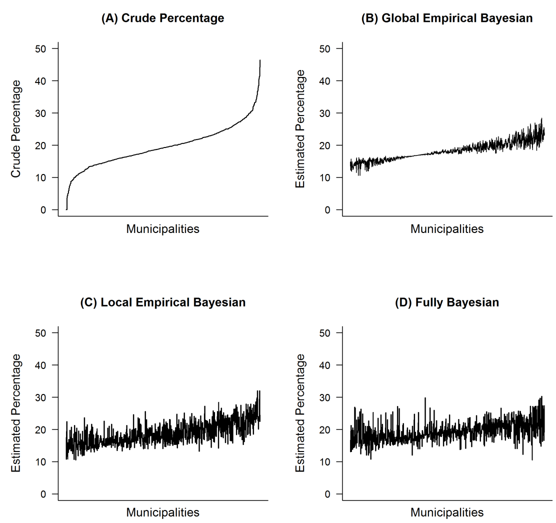

Figure 5 - Comparison between the (A) crude percentages of LBAM and the adjusted percentages of LBAM estimated from the (B) global empirical Bayesian (GEB) method, (C) local empirical Bayesian (LEB) method and (D) fully Bayesian (FB) model.

5 Discussion

The spatial statistical methods used in this study suggest that the percentages of LBAM are not randomly distributed across the State of Minas Gerais. The percentages of LBAM tend to be higher in the North than in the South of the State. This suggest an important association between the occurrence of adolescent pregnancy and socio-economic indicators, given that the North is the poorest region of the State while the South region concentrates the municipalities with the highest rates of development. These results are in accordance with those of Hau et al. (2009), Nogueira et al. (2009), Martinez et al. (2011) and Martins et al. (2014).

The results found in the present study by using estimation methods have shown that smoothing is necessary to better understand the pattern of the LBAM percentage in the State of Minas Gerais. The local and global empirical Bayesian methods were adjusted by means of the software TerraView, which has the advantage of being easy to use, although no confidence interval is generated for amounts of interest or covariable adjustments. On the other hand, the fully Bayesian method, fitted by using the software WinBUGS, generates credible intervals and allows the definition of a prior probability distributions and joint distribution of parameters of interest, thus increasing the knowledge on the precision of the estimates. However, the software WinBUGS brings obstacles to researchers who have little knowledge about programming languages. Computationally, the results obtained from both empirical Bayesian and fully Bayesian methods are alike, as observed in other studies using similar tools (SILVA et al., 2011; BERNADINELLI & MONTONOLLI, 1992). However, the latter allows statistical inferences to be made, which is more advantageous.

Acknowledgments

The authors would like to thank Brazilian research-funding agencies CAPES (Process number 4576/2014-04) and CNPq (Process number 305942/2012-3).

ROZA, D. L.; MARTINEZ, E. Z. Métodos empíricos e totalmente bayesianos para a identificação da distribuição espacial da gravidez adolescente em um estado brasileiro. *Rev. Bras. Biom.,* São Paulo, v.34, n.2, p.243-257, 2016.

- RESUMO: Métodos estatísticos espaciais são frequentemente utilizados para descrever a distribuição espacial da ocorrência de um evento de interesse em um espaço geográfico. No presente artigo, comparamos os métodos bayesianos global e local e o método totalmente bayesiano para identificar a distribuição espacial das porcentagens de nascidos vivos de mães adolescentes em Minas Gerais, o maior estado do sudeste do Brasil. O método totalmente bayesiano considera uma distribuição condicional autorregressiva (CAR) para os componentes espaciais do modelo. Os métodos usados neste estudo sugerem que as porcentagens de nascidos vivos de mães adolescentes não são aleatoriamente distribuídas no estado. Estas porcentagens tendem a ser mais altas na região norte que na região sul do estado. Isto sugere uma importante associação entre a ocorrência de gravidez adolescente e indicadores socioeconômicos, dado que o norte é a região mais pobre do estado enquanto a região sul concentra os municípios com as maiores taxas de desenvolvimento.

- PALAVRAS-CHAVE: Métodos estatísticos espaciais; métodos bayesianos, gravidez adolescente, estudo ecológico.

References

AGUERO-VALVERDE, J.; JOVANIS, P. P. Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis & Prevention*, Oxford, v.38, n.3, p.618-625, 2006.

ASSUNÇÃO, R. M.; BARRETO, S. M.; GUERRA, H. L.; SAKURAI, E. Maps of epidemiological rates: a Bayesian approach. *Cadernos de Saúde Pública*, Rio de Janeiro, v.14, n.4, p.713-723, 1998

BAILEY, T. C. Spatial statistical methods in health. *Cadernos de Saúde Pública*, Rio de Janeiro, v.17, n.5, p.1083-1098, 2001.

BERNADINELLI, L.; MONTONOLLI, C. Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statistics in Medicine*, Chichester, v.11, n.8, p.983-1007, 1992.

BESAG, J.; KOOPERBERG, C. L. On conditional and intrinsic autoregressions. *Biometrika*, London, v.82, n.4, p.733-746, 1995.

CARVALHO, M. S.; SOUZA-SANTOS, R. Analysis of spatial data in public health: methods, problems, and perspectives. *Cadernos de Saúde Pública*, Rio de Janeiro, v.21, n.2, p.361-378, 2005.

CASTRO, M. S. M.; VIEIRA, V. A. Space-time analysis of lung cancer mortality in Southern Brazil. *Revista Brasileira de Epidemiologia*, vol.7, n.2, p.131-43, 2004.

ELLIOTT, P.; WARTENBERG, D. Spatial epidemiology: current approaches and future challenges. *Environmental Health Perspective*, Durham, v.112, n.9, p.998-1006, 2004.

HAU, L.C.; NASCIMENTO, L. F. C.; TOMAZINI, J.E. Geoprocessing to identify the pattern of birth profile in Vale do Paraíba. *Revista Brasileira de Ginecologia e Obstetrícia*, Rio de Janeiro, v.31, n.4, p.171-176, 2009.

ILLIAN, J.; PENTTINEN, A.; STOYAN, H.; STOYAN, D. *Statistical analysis and modelling of spatial point patterns*. Chichester: John Wiley & Sons, 2008. 560p.

LAST, J. M. *A Dictionary of Epidemiology*. 4th ed. Oxford: Oxford University Press, 2001. 224p.

LAWSON, A. B. *Statistical methods in spatial epidemiology*. Chichester: John Wiley & Sons, 2013. 424p.

LAWSON, A. B.; KLEINMAN, K. *Spatial and syndromic surveillance for public health*. Chichester: John Wiley & Sons, 2005. 269p.

LUNN, D. J.; THOMAS, A., BEST, N.; SPIEGELHALTER D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, London, v.10, n.1, p.325-337, 2000.

MARSHALL, R. J. Mapping disease and mortality rates using empirical Bayes estimators. *Applied Statistics*, Oxford, v.40, n.2, p.283-294, 1991.

MARTINEZ, E. Z.; ACHCAR, J. A. Trends in epidemiology in the 21st century: time to adopt Bayesian methods. *Cadernos de Saúde Pública*, Rio de Janeiro, 2014, 2014, 30(4), 703-714.

MARTINEZ, E. Z.; ROZA, D.L.; CACCIA-BAVA, M. C. G. G.; ACHCAR, J. A.; DAL-FABBRO, A. L. Teenage pregnancy rates and socioeconomic characteristics of municipalities in São Paulo State, Southeast Brazil: a spatial analysis. *Cadernos de Saúde Pública*, Rio de Janeiro, v.27, n.5, p.855-867, 2011.

MARTINS, P. C. R.; PONTES, E. R. J. C.; PARANHOS FILHO, A. C.; RIBEIRO, A. A. Adolescent pregnancy: an ecological study in the health micro-regions of the State of Mato Grosso do Sul, Brazil – 2008. *Epidemiologia e Serviços de Saúde*, Brasília, v.23, n.1, p.91-100, 2014.

NOGUEIRA, M. J.; SILVA, B. F. A.; BARCELOS, S. M.; SCHALL, V. T. Analysis of the spatial distribution of adolescent pregnancy in the city of Belo Horizonte. *Revista Brasileira de Epidemiologia*, São Paulo, v.12, n.3, p.297-312, 2009.

SILVA, S. L. C.; FACHEL, J. M. G.; KATO, S. K.; BASSANESI, S. L. Patterns of variation in the infant mortality rate in Rio Grande do Sul State, Brazil: comparison of empirical Bayesian and fully Bayesian approaches. *Cadernos de Saúde Pública*, Rio de Janeiro, v.27, n.7, p.1423-1432, 2011.

SPIEGELHALTER, D. J.; BEST, N. G,.; CARLIN, B. P.; VAN DER LINDE, A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society - Series B.*, Oxford, v.64, n.4, p.583-616, 2002.

THOMAS, A.; BEST, N.; LUNN, D.; ARNOLD, R.; SPIEGELHALTER, D. *GeoBUGS user manual.* Cambridge: Medical Research Council Biostatistics Unit, 2004.

WALLER, L. A.; GOTWAY, C. A. *Applied spatial statistics for public health data.* Hoboken: John Wiley & Sons, 2004. 520p.

Received in 29.06.2015.

Approved after revised in 20.01.2016.

Appendix

WinBUGS code to the FB model used in the article:

```
model
{
for (i in 1:853)
{
y[i] ~ dbin(theta[i],N[i])
logit(theta[i]) <- alpha + gamma[i] }
alpha ~ dflat()
gamma[1:853] ~ car.normal(adj[], weights[], num[], tau.gamma)
tau.gamma ~ dgamma(.5,.0005)
sigma <- 1/tau.gamma
for (j in 1:sumNumNeigh) { weights[j] <- 1 }
}

list(num = c(3, 6, 9, 5, 11, 9, 2, 5, 5, 8, 4, 8, 4, 2, 5, 10,
4, 7, 4, ··· 6, 3, 3),
adj = c(497, 262, 212, 626, 607, 467, 172, 543, ··· 368, 234),
N = c(87, 220, 147, 48, 102, 189, 13, 32, 309 ··· 486, 49, 28),
y = c(13, 41, 23, 11, 20, 50, 2, 8, 67, 6, 69, ··· 99, 11,3),
sumNumNeigh = 4860 )
```