Bayesian model and spatial analysis of oral and oropharynx cancer mortality in Minas Gerais, Brazil

Modelo Bayesiano e análise espacial da mortalidade por câncer de boca e orofaringe em Minas Gerais, Brasil

Emílio Prado da Fonseca 1
Cláudia Di Lorenzo Oliveira 2
Francisco Chiaravalloti Neto 3
Antonio Carlos Pereira 1
Silvia Amélia Scudeler Vedovello 4
Marcelo de Castro Meneghim 1

Abstract The objective of this study was to determine of oral and oropharynx cancer mortality rate and the results were analyzed by applying the Spatial Analysis of Empirical Bayesian Model. To this end, we used the information contained in the International Classification of Diseases (ICD-10), Chapter II, Category C00 to C14 and Brazilian Mortality Information System (SIM) of Minas Gerais State. Descriptive statistics were observed and the gross rate of mortality was calculated for each municipality. Then Empirical Bayesian estimators were applied. The results showed that, in 2012, in the state of Minas Gerais, were registered 769 deaths of patients with cancer of oral and oropharynx, with 607 (78.96%) men and 162 (21.04%) women. There was a wide variation in spatial distribution of crude mortality rate and were identified agglomeration in the South, Central and North more accurately by Bayesian Estimator Global and Local Model. Through Bayesian models was possible to map the spatial clustering of deaths from oral cancer more accurately, and with the application of the method of spatial epidemiology, it was possible to obtain more accurate results and provide subsidies to reduce the number of deaths from this type of cancer.

Key words Oral cancer, Mortality, Epidemiology, Bayesian Analysis

Resumo O objetivo deste estudo foi determinar a taxa de mortalidade por câncer de boca e aplicar o Modelo Bayesiano Empírico e a Análise Espacial para suavizar a taxa bruta de mortalidade por câncer de boca e orofaringe. Foi usado o Capítulo II da Classificação Internacional de Doenças (CID-10) para Categorias C00 a C14. Os dados foram extraídos do Sistema de Informações sobre Mortalidade do Brasil (SIM) do Estado de Minas Gerais. As estatísticas descritivas e as taxas brutas de mortalidade foram calculadas para cada município. Posteriormente, foram aplicados os estimadores Bayesinos Empíricos. Em 2012, em Minas Gerais, foram registradas 769 mortes de pacientes com câncer de boca e orofaringe, com 607 (78,96%) homens e 162 (21,04%) mulheres. Houve uma grande variação na distribuição espacial da taxa de mortalidade bruta. Foi possível identificar aglomeração nas regiões Sul, Central e Norte mais precisamente pelo Estimador Global e Local. Através dos modelos Bayesians foi possível mapear aglomeração espacial de mortes por câncer de boca e orofaringe com maior precisão. Assim, pela epidemiologia espacial foi possível obter resultados mais precisos e fornecer subsídios para ações para reduzir a mortalidade por este tipo de câncer.

Palavras-chave Câncer oral, Mortalidade, Epidemiologia, Análise Bayesiana
Introduction

Epidemiological studies of incidence of and mortality from oral cancer have demonstrated high rates for this type of neoplasm in individuals, irrespective of age, sex, race and country. Regional differences in the distribution of mortality are attributed to the population’s exposure to risk factors, such as lifestyle, socioeconomic conditions, migratory phenomena, type of diet, or those directly related to the genetic characteristics of the population.

Brazil is among the countries with the highest incidence of and mortality from oral cancer. The estimates of incidence rates are 10.6 new cases per 100 thousand men and 3.4 new cases per 100 thousand women. Higher crude rates of mortality for men are also expected.

Studies published in the literature invariably analyze the data of mortality from oral cancer by crude or pondered rates by age or sex. However, this measure may be equal to zero or close to zero in municipalities with small populations, and hardly any or no cases. This result arises from analysis performed from crude rates, in which the information about neighboring municipalities are not included in calculation of the rate. Furthermore, spatial analysis also allows for identifying the formation of agglomerates of deaths and their relations with contextual and environmental factors.

Whereas, the use of the Empirical Bayesian method before the construction of thematic maps, refines the mortality rates, since they have been used in studies of other types of cancer. This refinement of the mortality rates results from the fact that the Empirical Bayesian method corrects random fluctuations, especially those related to locations with small populations and allows for better control of the variability of rates through the reduction in variance, considering information of the neighborhood and relations between close areas in small populations.

Considering as null hypothesis that mapping of the crude rate does not clearly identify the formation of agglomerates of deaths, and that the Empirical Bayesian analysis provides a re-configuration of the distribution of mortality and formation of agglomerates, the aim of this study was to compare the results obtained between the estimate by the crude rate and the Bayesian method for the rate of deaths from oral and oropharynx cancer, in the State of Minas Gerais, in 2012.

Methodology

The State of Minas Gerais is located in the South-eastern region of Brazil. It consists of 853 municipalities and has a population of 19,597,330 inhabitants (10% of the population of the country). It has a Mean Human Development Index (MHDI) of 0.731 and GINI Index of 0.46. The proportion of persons resident in urban areas represents 85.3%, and the economically active population is 9.94 million. It has a tropical climate with mean annual temperatures of over 18°C.

This was an ecological study with a secondary source of data. To analyze the situation of oral and oropharynx cancer in Minas Gerais, data made available by the System of Information about Mortality (SIM). Data collection was standardized by the selection of death per place of residence. To verify the consistency of information, the option was to access the data by means of two sites: State Secretary of Health of Minas Gerais (SES/MG) and Tabnet/DATASUS. This procedure was performed at two distinct time intervals to ensure that there would be no inclusion or exclusion of registers, and to guarantee the data, since the two systems are fed in a constant and consistent manner. Afterwards the data extracted from the system were tabulated in a computer software program. The data contained in the SIM are made available in a public and open manner, and as it was a secondary source of data. We use the year 2012 because while collecting data from the years 2013, 2014 and 2015 were not updated by the Mortality Information System (SIM) and can be modified in order to perform additional tests to confirm the origin of the lesion and cause of death.

The selection of locations of neoplasm was based on Chapter 2 of CID-10 (List of Tabulation CID-BR-10) for the year 2010. For this study, the following categories were selected: C00 to C14. The consistency and validation of the use of data extracted from SIM as source of information for studies with reference to mortality from oral and oropharynx cancer were performed by previous studies.

Crude Rate

The crude rate was calculated by the number of cases of each geographical unit divided by the population at risk and later multiplication by a base of 100 thousand persons.
Spatial Rate, Empirical Bayesian and Spatial Empirical Bayesian Models

The Crude Rate of mortality from cancer of the mouth for the year 2012 was adopted, as standard for comparison for the application of smoothing by the Bayesian models\textsuperscript{10,11,14}.

First was calculated the neighborhood matrix with contiguity criterion. The Spatial Rate is estimated by substituting the rate of each municipality with the rate of the region, and inserting the municipality itself and its neighbors in the calculation\textsuperscript{10,11,14}. The Empirical Bayesian Rate is the pondered sum between the Crude Rate and the overall/global mean rate. Ponderation has a factor inversely proportional to that of the population exposed; that is, the higher the population of a municipality, the greater its reliability. Therefore, the Empirical Bayesian Rate is close to the Crude Rate. In the literature, it is also called the Global Empirical Bayesian rate\textsuperscript{10,11,14}. The difference from the Spatial Empirical Bayesian Rate is that in the latter, ponderation is done by the local mean value, and is also called the Local Empirical Bayesian Rate\textsuperscript{10,11,14}. The methodology of this study was based on that of a previous study developed by Carvalho et al.\textsuperscript{12}, who used the Bayesian rates for mapping homicides in Brazilian municipalities.

In order to describe the geographical pattern of the occurrence of oral cancer, thematic maps were constructed, based on the distributions of the mortality rates for each municipality. The legends of the maps were standardized into five extracts to facilitate visualization and understanding of the results. For the construction of the maps, an electronic cartographic base was used, made available by IBGE, and a specific computer software program\textsuperscript{24,25}.

In addition, to investigate the relationship between age and gender were calculated the reason prevalence and odds ratio. The data were subdivided into two variables: gender (male and female) and age (less than 60 years and greater than 60 years).

Results

The distribution of deaths from oral and oropharyngeal cancer of the 853 municipalities of the State of Minas Gerais in 2012 were analyzed. In the mentioned year, 770 deaths from cancer of the mouth occurred, with 608 (78.96%) being men and 162 (21.04%) women. For this study, one death was excluded because of lack of information about the municipality of residence. Considering the estimated population of the State in 2012\textsuperscript{15} and the total number of cases, the crude rate for Minas Gerais in 2012 was 3.87 deaths per one hundred thousand inhabitants. The proportion of male deaths was 3.75 higher than the proportion of women deaths.

The simple mean of the Crude Rate and Spatial Empirical Bayesian Rate of deaths from cancer of the mouth were lower, in addition to presenting a higher standard deviation in comparison with the Empirical Bayesian and Spatial Rates (Table 1).

In 2012, 517 municipalities in Minas Gerais recorded no deaths; In the same year, 176 recorded only 1 death. Among the 8 less populated municipalities, and with few cases, the crude rate ranged from 0 to 36.66 and caused an overestimation of the death rate from this type of cancer. In an inverse direction, the 4 most populated municipalities, and with the highest number of deaths in the state, presented crude rates that ranged from 2.58 to 5.38 and caused an underestimation of the rate (Table 2). In Minas Gerais, men over 60 years old die 1.38 times more than women with over 60 years old when compared to deaths in individuals less 60 years old. The prevalence of deaths from oral and oropharynx cancer was 3.75 more in men than in women (Table 3).

Figures 1 and 2 present the distribution of the mortality rates from oral and oropharynx cancer that occurred in the state in 2012. From Figure 1, it is possible to affirm that deaths from cancer of the mouth are dispersed throughout the State and interspersed by a high number of geographic voids and low extremes represented by lighter colors. As the method used was the crude rate, it is also possible to visualize that the highest rates occurred in municipalities of small size, represented by the darker colors.

In Figure 2, the redistribution of the crude rate proposed by the Bayesian and Spatial rates is pointed out; that is, predominance of values represented by the intermediate layers in all the directions of the state. There was a reduction in the values of the lower layer, and increase in the darker layer. This may be as a result of using both the global mean and local mean for calculating the rates. It is possible to observe that the Spatial Rate behaved in a manner similar to that of the Spatial Empirical Bayesian Rate, so that, the map of the Spatial Empirical Bayesian Rate is smoother than the Spatial Rate. This may have occurred because both are related to the local mean value.
It is also possible to observe the formation of regional agglomerates of deaths. The Empirical Bayesian Rate presented greater scattering among the segments, mosaic appearance and many polygons. This may be due to the influence of the general mean which was 3.87 deaths per 100 thousand inhabitants. The large number of small municipalities with zero cases may contribute to the downward displacement of the global mean value and may consequently have compromised this estimate.

**Discussion**

In this study, techniques for smoothing of data for mortality rates from oral and oropharynx cancer were used. For rare events, and in the case...
of the occurrence of cancer, the researcher needs to be cautious, because problems of under- or overestimation of the rates found may occur\textsuperscript{23}. In municipalities in which no cases occurred, or which have very small populations, the crude rate would be zero or close to zero, so that the appearance of a single case in these municipalities may significantly alter the crude rate\textsuperscript{23}. To smooth the analysis, the Bayesian models and identification of agglomerates of death from oral and oropharynx cancer were used in a manner similar to those of previous studies that used Bayesian models for stabilizing mortality rates from oral cancer and identification of agglomerates\textsuperscript{26,27}.

Researches on oral and oropharynx cancer in Brazil have used crude and pondered rates\textsuperscript{28-30}. However, they are sensitive to random variability, particularly in municipalities with low populations. In this study, the phenomenon of overestimation and underestimation of the crude rate were demonstrated. This may occur in other studies on oral and oropharynx cancer, which are based on frequency, crude and pondered rates. In the state studied, a municipality with over 2 million inhabitants and 120 deaths recorded a lower crude rate than a municipality with 2,054 inhabitants and 1 death. It could be verified that the Bayesian and Spatial rates smoothed the results of crude rates. That is, as many municipalities in the state did not register any deaths from this type of cancer, and had a crude rate equal to zero, the mean value of the crude rate was displaced by lower values. These facts compromised the estimation of the crude rate and showed the influence of population size on the value of the results\textsuperscript{23} (Tables 1 and 2).

Identification of agglomerates of death from oral cancer has also been made by means of frequencies and pondered rates by age or gender\textsuperscript{3}. The results of this study suggest the Bayesian models as alternative to the use of crude, pondered and frequency rates for analysis, mapping and identification of agglomerates of death from this type of cancer that occurred in Minas Gerais.

Therefore, the methods used suggest the smoothing of extreme values by means of Bayesian and Spatial techniques. For the more populated municipalities, the corrected rates obtained
values with little variation in comparison with the crude rate. Whereas for smaller municipalities, the smoothing converges on values close to those of the crude rates in the larger municipalities. This result was similar to that found in a study conducted in Brazil on homicide rates\(^{23}\) (Figures 1 and 2).

In a state with many municipalities, regional differences and the high number of municipalities that did not present deaths, use of the Empirical Bayesian Rate (EBR) is not recommended, because its correction is made by the mean value of the state. Among the rates corrected, only the Spatial Empirical Bayesian Rate (SEBR) presented better visualization of the distribution of deaths and identification of the regional agglomerates (Figure 2). Smoothing by SEBR uses the local mean value, the municipalities located in microregions have the same characteristics, with coherent regional results. Therefore, SEBR may be a more reliable measure of smoothing for studies on mortality from oral and oropharynx cancer in Minas Gerais.

By means of both the Spatial and Empirical Rates (Global and Local) it was possible to identify agglomerates of deaths in regions with the best and worst socioeconomic conditions in the state (Figure 2). A previous study supposed that the problem was not related to the size of population or level of development, because without distinction among regions, advanced cases of the disease, at a high level of complexity occurred on arrival, therefore, with an unfavorable prognosis at the beginning of treatment\(^7\). However, previous studies have identified the higher frequency of deaths from oral cancer in regions with the

---

**Figure 2.** Spatial distribution of the mortality rates from oral and oropharynx cancer, adjusted by the Bayesian models (per 100 thousand inhabitants), MG, 2012.
worst socioeconomic indicators\textsuperscript{3,4,7,30}. Epidemiological studies about oral and oropharyngeal cancer prevalence and association with contextual, environmental and individual variables are based on frequencies and rates\textsuperscript{5-7,13,27-31}. It is important to affirm that the Bayesian rates do not substitute the crude and pondered rates, but make a smoothing\textsuperscript{26,27}. This study suggests that the maps and identification of agglomerates of deaths from oral and oropharynx cancer should be conducted\textsuperscript{10,11,14,26,27}.

It is important to point out that this method is subject to error caused by under-registration, particularly in small municipalities, and when the quality of the information is not known. Under-registration of deaths and quality of information present in the declaration of Death may compromise the quality of SIM and explain the high number of municipalities with “zero” deaths from cancer of the mouth and oropharynx in Minas Gerais\textsuperscript{21,22}. Therefore, epidemiological studies of spatial distribution based on maps are sensitive to the quality of the data of the areas studied.

This study has limitations. For being a cross-sectional study could not establish causality. By using secondary data may have occurred bias information.

Conclusion

The application of smoothing of crude rates allowed a alternative interpretation of the results. However, this application suffers influence from the size of the population of the municipality. There was the formation of agglomerates in certain regions, and this fact provides subsidies for the development of priority actions for facing up to this type of death.

Collaborations

EP Fonseca, DL Oliveira, F Chiaravalloti Neto participated in the design and development of the article, collection and tabulation of data, data interpretation and writing of the article. AC Pereira and SAS Vedovello attended the article design, data interpretation and critical text revision and approval of the final version. MC Meneghim attended data interpretation, critical text revision and approval of the final version.

Acknowledgements

The Authors thank INPE for making available the computer software program TerraView ® free of charge, for the construction of thematic maps.
References