
Stomach cancer incidence in Southern Portugal 1998-2006: a spatio-temporal analysis ^a

A.L. Papoila^{1,*}, C.Ribeiro², R. São João³, C.Geraldes¹, A. Amaral Turkman⁴ and A. Miranda⁵

^aResearch partially sponsored by national funds through the Fundação Nacional para a Ciência e Tecnologia, Portugal - FCT under the projects PEst-OE/MAT/UI0006/2011 and PTDC/MAT/118335/2010.

¹ Faculty of Medical Sciences, UNL/CEAUL, Lisbon, Portugal; ana.papoila@fcm.unl.pt, cgeraldes@fcm.unl.pt

² ISE-UALG/CEAUL, Algarve, Portugal; cribeiro@ualg.pt

³ Instituto Politécnico de Santarém/CEAUL, Santarém, Portugal; ricardo.sjoao@esg.ipsantarem.pt

⁴ Faculty of Sciences, University of Lisbon/CEAUL, Lisbon, Portugal; maturkman@fc.ul.pt

⁵ Registo Oncológico Regional Sul, Lisbon, Portugal; amiranda@ipolisboa.min-saude.pt

*Faculty of Medical Sciences, UNL/CEAUL, Lisbon, Portugal; ana.papoila@fcm.unl.pt

Abstract. *Understanding geographical differences in health, particularly in small areas, became a major concern of epidemiologists. Geographical association studies and, more recently, several spatial disease mapping studies have emerged due to the development of new spatial statistical tools. Among other diseases, these methods are being applied to analyze cancer data. However, in this kind of studies, it is of utmost importance to also investigate the influence of temporal variability and that is why spatio-temporal studies became so popular. The aim of this study is to investigate spatial and temporal trends for the incidence of this type of cancer. This retrospective population-based study is based on data on all stomach cancers registered by the Southern Portuguese Cancer Registry (ROR Sul) between 1998 and 2006. Because several studies have underlined the important role of socioeconomic status in cancer risk, information on this variable has also been taken into account. Bayesian hierarchical models were applied to model stomach incidence at a county level and resulting relative risks were used to build risk maps for cancer incidence. Age-Period-Cohort models were also applied.*

Keywords. *Spatio-temporal analysis; Bayesian hierarchical models; INLA; Age-period-cohort models; Stomach cancer.*

1 Introduction

Stomach cancer is the fourth most common cancer and the second leading cause of cancer-related death worldwide. In cancer epidemiology, studies that lead to an understanding of factors that originate cancer incidence trends are of the utmost importance. Age-Period-Cohort (APC) models [1] allow the study of the effect of patients age at diagnosis (age), date of diagnosis (period) and date of birth (cohort) on cancer incidence. Bayesian Hierarchical (BH) models [2] allow the incorporation of spatial and temporal effects through prior information and enable ecological analyses of associations between cancer incidence and potential risk factors over aggregated areas. The knowledge of tendencies in cancer incidence rates estimated by APC and BH models are an additional and important tool in the definition of policies in oncology. Based on official cancer registries of Southern Portugal (ROR-Sul), this study aims to model the incidence rates of stomach cancer during the period 1998-2006, using APC and BH models.

2 Materials and Methods

This population-based study considered 9770 diagnosed cases in 1998-2006, registered in ROR-Sul (109 counties), of which, 58.6% were men.

Let Y_{ijk} be the observed number of cancer cases among n_{ijk} individuals, at risk, in the population of county i , diagnosed with stomach cancer in year j and of gender k . A Poisson approximation of the binomial distribution allows us to consider $Y_{ijk} \sim \text{Poisson}(E_{ijk}\theta_{ijk})$, where E_{ijk} is the expected number of the population at risk for the disease and θ_{ijk} is the relative risk. Because cancer is a disease that affects individuals of different age classes disproportionately, standardized rates must be calculated. Considering eighteen 5-year age classes, let y_{ijkl} be the observed number of cases and n_{ijkl} the observed number of the population at risk, for individuals in age class l , ($l = 1, \dots, 18$) and gender k , ($k = 1, 2$), living in county i , ($i = 1, \dots, 109$) and diagnosed with cancer in year j , ($j = 1, \dots, 9$). The corresponding expected number of cases is given by

$$E_{ijk} = \sum_l n_{ijkl} \frac{\sum_i \sum_j \sum_k y_{ijkl}}{\sum_i \sum_j \sum_k n_{ijkl}}.$$

The following spatio-temporal model for the relative risk was assumed

$$\log(\theta_{ijk}) = \beta_0 + \beta_1 \text{gender}_k + \beta_2 \text{socio.economic.status}_i + \psi_i + v_i + (\gamma + \delta_i)t_j,$$

where β_0 is the intercept, β_1 and β_2 are fixed effects, ψ_i are random effects accounting for spatial heterogeneity and v_i are random effects accounting for unstructured heterogeneity. We also assume that ψ_i and v_i are mutually independent, with conditionally autoregressive (CAR) and $Normal(0, \sigma_v^2)$ priors, respectively; γt_j is a linear trend term in time t_j , δ_i is an interaction random effect between space and time, with a CAR prior. We also assigned the following independent diffuse priors for the hyperparameters: $\beta_0, \beta_1, \beta_2, \gamma \sim Normal(0, 1000)$, $(\sigma_\psi^2)^{-1}, (\sigma_v^2)^{-1}, (\sigma_\delta^2)^{-1} \sim Gamma(1, 0.01)$.

The above model was fitted and, because interaction was not detected and counties' socioeconomic status was not significant, they were excluded from the model. On the other hand, gender emerged as an important factor in explaining stomach cancer incidence. In fact, it is well known that gender is a relevant variable for all types of cancers and, because there is a suspicion of a different spatio-temporal behavior, two BH models were used, one for males and another one for females. For model comparison, the Deviance Information Criteria and the cross-validated log score were used. BH models were implemented using R-INLA [2]. Concerning APC models, they were implemented using generalized additive models with a Poisson response probability distribution, where age, period and cohort effects were modelled by smooth functions [1]. They were also implemented using R.

3 Results

Results from APC models show that, according to age, both genders have a similar increase in incidence, with evident higher rates for male gender (black line) for ages over 50 years. The period seems to have no impact and a cohort effect is highly visible, showing a decrease in rate ratios for birth cohorts born after 1948 (median cohort).

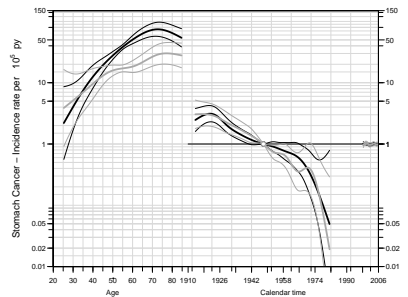


Figure 1: APC model estimated effects of age, cohort and period, using 1948 cohort as reference.

Results from BH model showed that males are at a higher risk (Bayesian regression coefficient estimate: 0.77; 95% credible interval (CI):(0.73, 0.81). Concerning time, a significant decrease in cancer incidence was found (Bayesian regression coefficient estimate: -0.04 ; 95% CI:(-0.05 , -0.03). Both spatially estimated variances, structured and unstructured, were low ($E(\sigma_{\psi}^2|data) = 0.022$, 95% CI:(0.005, 0.066); $E(\sigma_v^2|data) = 0.019$, 95% CI:(0.008, 0.038), respectively). Results from the models fitted to each gender

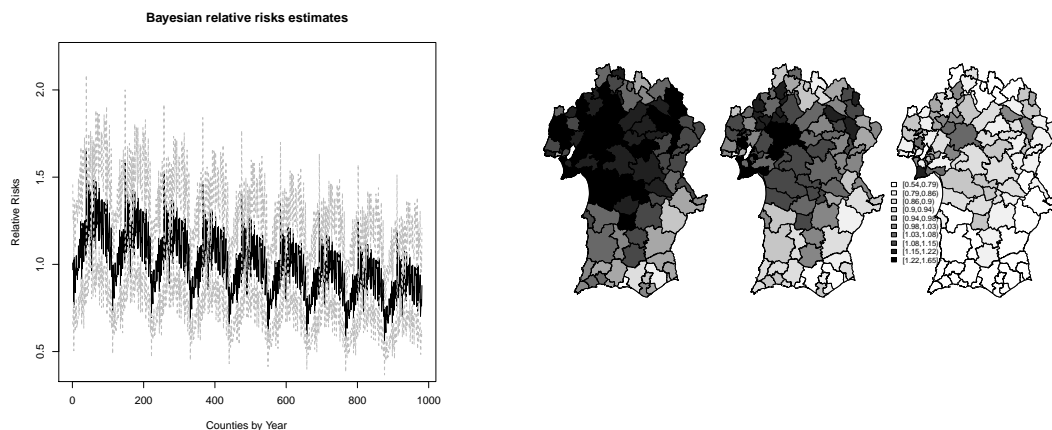


Figure 2: Female gender: posterior expected relative risks with 0.025 and 0.975 quantiles (left); risk maps for cancer incidence in 1998, 2001 and 2006, respectively (right). The estimates of the relative risks attained a minimum of 0.55, a maximum of 1.65 and a median of 0.98 ($Q_{0.25} = 0.88$, $Q_{0.75} = 1.11$).

separately showed that, although low values for the estimates of the variance persisted, male and female gender have different spatial correlation patterns. In fact, for females, the variance of the spatially structured component was higher than the variance of the spatially unstructured component ($E(\sigma_{\psi}^2|data) = 0.042$, 95% CI:(0.009, 0.120); $E(\sigma_v^2|data) = 0.015$, 95% CI:(0.004, 0.042), respectively); for males,

the opposite happened ($E(\sigma_{\psi}^2|data) = 0.014$, 95% CI:(0.003,0.048); $E(\sigma_{\nu}^2|data) = 0.020$, 95% CI:(0.008,0.038), respectively). This difference is consistent with Figures 2 and 3 (left) where a clear pattern of the female relative risks of all counties in each year is shown. The decreasing trend of the incidence risks along the study period is visible at Figures 2 and 3; time coefficient estimates and 95% CIs were: $-0.04;(-0.05, -0.03)$ and $-0.06;(-0.07, -0.05)$ for female and male genders, respectively.

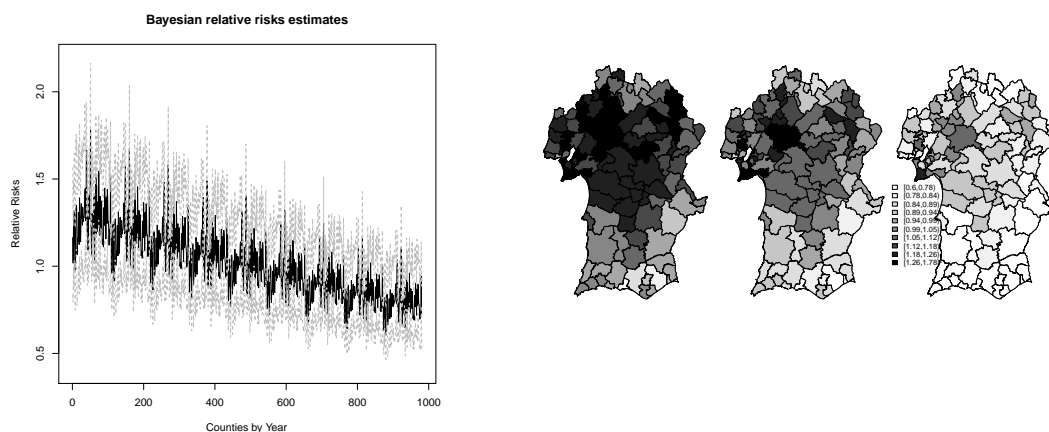


Figure 3: Male gender: posterior expected relative risks with 0.025 and 0.975 quantiles (left); risk maps for cancer incidence in 1998, 2001 and 2006, respectively (right). The estimates of the relative risks attained a minimum of 0.61, a maximum of 1.78 and a median of 0.99 ($Q_{0.25} = 0.87, Q_{0.75} = 1.14$).

4 Conclusions and Discussion

APC models results showed that age and cohort have a strong effect on the incidence of stomach cancer. BH models results showed that region influence on cancer incidence was weak, particularly for men. A decrease in incidence over the study period, in a more pronounced way for males, was found. Concerning this decrease, results from APC and BH models seem to be contradictory, however, because age and cohort were not considered in BH models as covariates and because $cohort = period - age$, cohort effect was transferred to the period effect in the BH models. The exact causes of stomach cancer still remain unknown, however, among other risk factors, changes in dietary and lifestyle habits as well as the control of *Helicobacter pylori* infection may be related to this decreasing trend.

Acknowledgments. The authors wish to thank Prof. Fátima Pina for making available the data concerning socioeconomic status of each county.

References

- [1] Carstensen, B. (2007). Age-period-cohort models for the Lexis diagram. *Statistics in Medicine*, **26**:3018–3045.
- [2] Rue, H., Martino, S. and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations, *Journal of the Royal Statistical Society B*, **71**(2), 319–392.