The longevity pattern in Emilia Romagna: a spatio-temporal analysis

Giulia Roli, Rossella Miglio, Rosella Rettaroli and Alessandra Samoggia

Abstract In this paper, we investigate the pattern of longevity in Emilia Romagna, a North-Eastern region of Italy, at a municipality level. We consider a modified version of Centenarian Rate in two different periods. Spatio-temporal modeling is used to tackle at both periods the random variations in the occurrence of long-lived individuals, due to the rareness of such events in small areas. This method allows to exploit the spatial proximity smoothing the observed data, as well as to control for the effects of a set of regressors. As a result, clusters of areas characterized by extreme indexes of longevity are well identified and the temporal evolution of the phenomenon can be depicted. In addition, we evaluate the effects of the structure of mortality on the cohort of long-lived subjects in the second period. A spatial analysis is carried out by including the territorial patterns of mortality in a longitudinal perspective. We control for the major causes of death in order to deepen the analysis of the observed geographical differences.

Key words: Aging, mortality and health; Spatial Data Analysis.

1 Introduction

In the last decades, the study of human longevity and its development has drawn the attention of researchers belonging to different fields of analysis. Various studies performed in different Italian regions showed the presence of specific areas where the prevalence of oldest-old people is higher than elsewhere. For instance, a definite geographical area in Sardinia is characterized by an exceptional male longevity ([9]), as well as a low female/male centenarian ratio; a significant negative correlation between surname abundance and index of longevity has been detected in Calabria

Giulia Roli, Rossella Miglio, Rosella Rettaroli, Alessandra Samoggia
Dipartimento di Scienze Statistiche, Università di Bologna.
e-mail: g.roli@unibo.it

where some isolated areas of male longevity present a high level of inbreeding ([7]); in Emilia Romagna some longevity "clusters" have been identified and their persistence has been detected by comparing the results of different spatial scan statistic methods ([8]).

The explanatory analysis of disparities in the frequencies of the oldest-old population reminds of different genetic and environmental factors differently spread at a geographical level. In order to deepen these several aspects, one of the scientific approaches aims at mapping the geographical diffusion and the temporal evolution of extreme longevity. With this purpose, a recent statistical approach consists in the use of spatial and spatio-temporal models (see, e.g., [6]) which involve the geographical proximity and the time interaction of the areas resulting in a smoothing effect among the observations. As a result, more consistent estimates and inferences on quantities of interest are provided. Moreover, these methods allow to control for extra-variation among the units ([4]). Indeed, when the geographical analysis is performed on a fine territorial scale and the phenomenon under study is characterized by a scarce number of units, the territorial distribution of the cases can be invalidated by random variations.

In this paper, we use a Bayesian spatio-temporal model ([1], [5]) to manage both the geographical structure and the temporal dimension of the extreme longevity. In particular, we consider a modified version of Centenarian Rate ([10]) in two different five-years periods (1995-1999 and 2005-2009), separately by municipality and gender. The small numbers at municipality level are properly tackled by the model, which further allows to embed into the analysis some areal features. The main purposes are to identify territorial groups of areas characterized by high or low levels of longevity and to investigate the development of these clusters by time.

Besides being influenced by spatial interactions and areal features, it is well known that the higher or lower presence of long-lived people across the areas largely depends on mortality features after 80 years ([11], [12]). Therefore, we consider the last fifteen years levels of mortality which have affected at younger ages the survival of the observed cohort of long-lived individuals. In particular, we aim at explaining how the current distribution of longevity is linked to the past mortality of the same cohort. Within this longitudinal framework, the longevity outcome is modeled by a Bayesian spatial regression where the area-specific levels of mortality for each group of age in the cohort are included as regressors. The hierarchical structure of the model further allows to take into account correlations among the mortality covariates.

The paper develops as follows. We firstly describe the regional area under study and corresponding data and indicators. Section 3 defines the statistical analysis. The results are summarized and discussed in Section 4. The last section reports the main conclusions.

2 Area, data and indicators

Emilia Romagna is a North-Eastern Italian region which shows one of the oldest age structures in Europe (with 22.5% persons aged 65+, and 6.9% persons aged 80+ in 2009). The region is characterized by a great geographical variability in terms of environmental context, social conditions and economic resources. Emilia Romagna is split up into nine provinces and 341 municipalities.

The different spread of longevity in the regional area is measured by a modified version of the Centenarian Rate (CR) ([10]). This indicator, denoted by CR^{95+}, is obtained at a municipality level by dividing the count of people aged 95 and over, P^{95+}, by the number of 55-64 years old persons living in the same area 40 years earlier, P^{55-64}, separately by gender. In this framework, we assume that the individuals resident in an area 40 years before the time point t under study are the population exposed to the "risk" of becoming P^{95+}. To depict the temporal dimension of the phenomenon, we calculate the CR^{95+} separately by sex and municipality area in two five-years periods: 1995-1999 and 2005-2009.

Additional information on the features of the municipalities are collected. In detail, we consider the classification of areas with respect to the altimetry zone and population density. These are then combined to form the following groups: mountain, hill and density lower than 78 people per km^2, hill and density between 78 and 198, hill and density greater than 198, coastal hill, plain and density lower than 78, plain and density between 78 and 198, plain and density greater than 198.

For the cohort of P^{95+} in the period 2005-2009, we focus on the corresponding patterns of mortality measured when they were aged 80-89 (in 1990-1994) and 90-99 (in 2000-2004). In particular, with respect to the municipalities and these age classes we compute the Standardized Mortality Ratios (SMR) for two groups of causes of death: malignant neoplasms and diseases of the circulatory system, which jointly represent the most common causes of death for people aged 80 and over. In order to facilitate the interpretation of the results, these SMRs are grouped into 3 levels of mortality based on their quintile distribution. The first and the last quintile subsets identify municipalities with extreme (low or high) mortality levels. Areas with values of SMR close to the overall regional rate are included into the residual group.

The sources of the data all refer to official statistics published by the Italian institutional agency for statistical data collection (ISTAT) and by the regional authorized agency of Emilia Romagna.

3 Methods

We develop two different models, separately by sex. The first investigates the space-time pattern of "risk" of becoming P^{95+} among the 341 municipalities of Emilia Romagna by adopting a hierarchical Bayesian approach which exploits the adjacency and interaction of the geographical areas. We further associate a temporal dimension to the phenomenon, by considering the evolution of P^{95+} in the two periods 1995-1999 and 2005-2009. In detail, with respect to the i-th municipality, we assume the observable P^{95+} at each time period t (1995-1999 and 2005-2009), denoted by y_{it}, are Poisson distributed with means $p_{it}\theta_{it}$. In this formulation, p_{it} represents the potential P^{95+} and θ_{it} is the estimate of CR^{95+} in the required location and period. Then, we follow the conventional log-linear formulation on the rate θ_{it} and we allow for the possibility of different components that additively contribute to explain the space-time distribution of these rates. We further control for the effects of some areal features by including the 8 categories of altimetry and population density introduced above. As a result, the model for the logarithm of the rate θ_{it} can be analytically expressed as follows:

$$log\,(\theta_{it}) = \alpha_t + \beta x_i + u_i + v_i \qquad (1)$$

where α_t represents the time-varying intercepts, x_i is the altimetry and population density group of the i-th municipality, β is the corresponding effect on the log rate, u_i and v_i are the correlated and uncorrelated spatial heterogeneity, respectively, which are both assumed to be constant in time.

For the correlated spatial component u_i, we assume a Gaussian conditionally autoregressive (CAR) model ([3], [1]). Although improper, the CAR prior leads to a posterior distribution which is proper, allowing the Bayesian inferences still proceed. The random effects v_i which capture the region-wide heterogeneity are supposed to follow an ordinary exchangeable normal prior. We specify vague and proper distributions for the other (hyper-) parameters.

In order to investigate the effects on longevity of the past structure of mortality by cause, we use a second spatial model which allows to adopt a longitudinal perspective. Indeed, it includes as regressors the SMRs for both malignant neoplasms and diseases of circulatory system that have characterized the survival of current (2005-2009) long-lived individuals when they were 90-99 years old (in 2000-04) and when they were 80-89 years old (in 1990-94). The outcome Y now represents the count of P^{95+} in years 2005 to 2009, lonely. For each municipality i, the y_i are again assumed to follow a Poisson distribution of parameter $p_i\theta_i$, with analogous meanings to those described in the previous section. The log-linear model on the rate θ_i involves the spatial structured (u_i) and unstructured (v_i) components, besides controlling for the effects altimetry and population density (x_i). The groups of SMRs by cause are denoted by z_{iad}, where a is the age group of the cohort and d is the specific cause of death. Thus, the model can be formalized as follows:

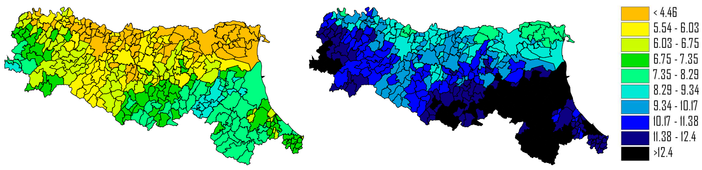
$$log\,(\theta_i) = \alpha_0 + \beta x_i + \sum_{a,d} \gamma_{ad} z_{iad} + u_i + v_i \qquad (2)$$

where the estimation of parameters γ_{ad} now represents our main objective.

In addition, we allow for a correlation among causes of death and among the repeated measure of mortality over time by modeling the priors on the effects γ_{ad} through a noninformative multivariate normal distribution with unknown population mean vector and a variance covariance matrix Σ.

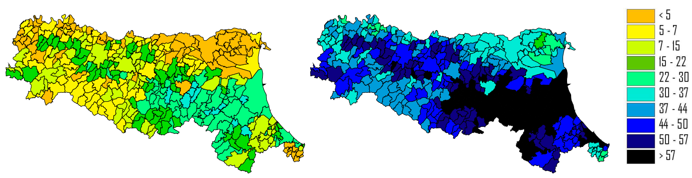
4 Results and discussion

The application of the hierarchical spatio-temporal model introduced above offers the opportunity to investigate different aspects of the longevity pattern. The comparison of the estimated CRs^{95+} across time shows an increase in the numbers of long-lived subjects, especially for women. The males are characterized by lower values with small differences among the areas. In order to evaluate the different territorial contributions to the CRs^{95+}, separately by sex, we further consider the ranking of the municipalities according to their decile distribution in both periods (Figures 1 and 2). All the results refer to a number of 1000 individuals exposed to "risk".



(a) 1995-1999. (b) 2005-2009.

Fig. 1 Estimated CRs^{95+}. Men.



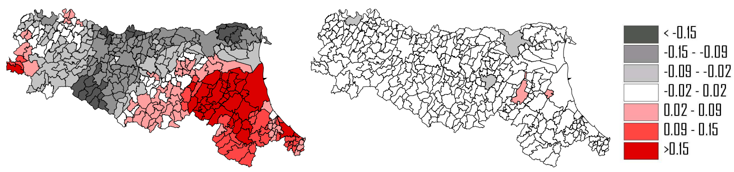
(a) 1995-1999. (b) 2005-2009.

Fig. 2 Estimated CRs^{95+}. Women.

There is a notable persistence of the areas of lower and higher occurrences of long-lived individuals, associated to the rise in the CRs^{95+} over time. These results show mean and median values higher than the regional ones in the municipalities belonging to the provinces of Ravenna and Forli-Cesena, spreading out in the Adriatic coast, at one side, and in the Apennine municipalities of Bologna and Modena, on the other. Especially for men, some areas of the province of Piacenza still stand out for high values of CR^{95+}. Conversely, the municipalities of the province of Ferrara are characterized by a lower longevity.

In addition, the two components of spatial variability, **u** and **v**, are evaluated. The first one detects the clustering effect of adjacent areas. For the male group, two big clusters with a positive contribution on the CRs^{95+} can be clearly identified (Figure 3 (a)). These substantially correspond to the same areas described before as characterized by the highest values of CR^{95+}. On the other hand, a negative effect is detected in some municipalities of Ferrara and Apennine areas of Parma and Reggio Emilia.

The second part of variability is due to the peculiarities of individual areas and results from municipalities with observed values significantly different from the other nearby areas. As a result, we identify some municipalities whose characteristics should be investigated in depth (Figure 3 (b)). For both sexes, the spatial variability due to clustering seems to prevail (78% of total variability for men and 67% for women).



(a) Clustering. (b) Heterogeneity.

Fig. 3 Spatial variability. Men

Table 1 Covariates.

	Men		Women	
	Fitted rate ratios	90% CI	Fitted rate ratios	90% CI
Mountain	1	-	1	-
Hill and density < 78	0.961	0.827 - 1.114	0.956	0.872 - 1.047
Hill and density $78 - 193$	0.978	0.846 - 1.127	1.132^a	1.037 - 1.233
Hill and density > 193	0.852	0.741 - 0.973	1.027	0.936 - 1.126
Coastal hill	0.895	0.620 - 1.295	0.751	0.590 - 0.952
Plain and density < 78	0.772	0.611 - 0.972	0.743^a	0.640 - 0.862
Plain and density $78 - 193$	0.794^a	0.690 - 0.912	0.942	0.859 - 1.033
Plain and density > 193	0.906	0.801 - 1.021	1.015	0.935 - 1.103

[a] Results significant at a 5% level.

The fitted rates for the intercepts in the two periods, $exp(\alpha_1)$ and $exp(\alpha_2)$, represents the level of probability of observing long-lived subjects in the mountain areas of Emilia Romagna, that is the reference group. An overall increase in these values can be observed over time (0.007 to 0.012 for men; 0.022 to 0.038 for women). For males, the other municipalities all experienced rate ratios which are lower than 1. Conversely, in the female group, people living in hill areas with medium density have a significant higher rate of becoming long-lived than mountain residents (Table 1).

Through the cohort spatial analysis we can further investigate the effects of the recent (age group 90-99) and past (80-89) structure of mortality on the corresponding cohort of long-lived individuals. We consider the period 2005 to 2009, lonely, when the larger territorial differences have been observed for both sexes.

Table 2 Effect of the past mortality on the cohort of P^{95+} in 2005-2009.

	Men		Women	
	Fitted rate ratios	90% CI	Fitted rate ratios	90% CI
Circulatory diseases (80-89):				
Medium mortality	1	-	1	-
Low mortality	1.091	0.996 - 1.195	1.005	0.947 - 1.067
High mortality	0.980	0.878 - 1.094	0.979	0.920 - 1.042
Circulatory diseases (90-99):				
Medium mortality	1	-	1	-
Low mortality	1.197[a]	1.081 - 1.326	1.068	1.009 - 1.130
High mortality	0.797[a]	0.714 - 0.914	0.848	0.795 - 0.906
Malignant neoplasms (80-89):				
Medium mortality	1	-	1	-
Low mortality	1.124	1.012 - 1.248	1.000	0.937 - 1.066
High mortality	0.828[a]	0.752 - 0.909	0.997	0.943 - 1.054
Malignant neoplasms (90-99):				
Medium mortality	1	-	1	-
Low mortality	1.070	0.963 - 1.190	1.036	0.974 - 1.100
High mortality	0.896	0.808 - 0.991	0.926	0.871 - 0.983

[a] Results significant at a 5% level.

As expected, high mortality levels in the SMRs yield rate ratios lower than 1 when compared with areas with medium values for both periods, causes of death and sexes; conversely, areas characterized by a lower intensity of mortality are more likely to have higher values of CR^{95+}. For both sexes, the mortality levels for the circulatory diseases between ages 90 and 99 seem to mostly affect the survival by 95 years. Moreover, a high mortality caused by malignant neoplasms at age 90 to 99 significantly reduces the consequent longevity. Furthermore, for men the chance of becoming long-lived considerably depends on the mortality levels between 80 and 89 years of age and, especially, on the deaths caused by malignant neoplasms. Conversely, female longevity seems to be weakly affected by the level of mortality in the first age group (i.e. 80-89) (Table 2).

5 Conclusions

The estimated values of the adopted longevity index allow to identify some groups of areas where the people reaching 95+ years of age are more likely to occur with a substantial persistence over time. As a result, the effect of some crucial patterns of mortality by age and cause, as well as a set of areal features and correlations, can be jointly investigated. The use of a hierarchical Bayesian modeling for spatial and spatio-temporal data to analyze the longevity pattern in small areas represents a quite innovative application, which aims to confirm the usefulness of these methods also in socio-demographic research.

A critical aspect of these methods is certainly represented by the sensitivity to the choice of the priors ([2]). Therefore, as a further development, a sensitivity analysis on different specifications of the priors will be carried out.

Since mortality and longevity may be differently affected by the same risk factors and the level of a specific-cause mortality might encourage/inhibit the presence of longevity over a region, another development of the analysis should consist in the use of a multivariate spatial model to properly analyze this kind of data under a cohort perspective.

References

1. Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). Hierarchical Modeling and Analysis for Spatial Data. London: Chapman and Hall.
2. Bernardinelli, L., Clayton, D. G., and Montomoli, C. (1995). Bayesian estimates of disease maps: how important are priors? Statist. Med. 14: 2411-2431.
3. Best, N.G., Waller, L.A., Thomas, A., Conlon, E.M. and Arnold, R.A. (1999). Bayesian models for spatially correlated diseases and exposure data. In Bayesian Statistics 6, eds. J.M. Bernardo et al. Oxford: Oxford University Press.
4. Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. Appl. Statist. 33(1): 33-44.
5. Lawson, A.B., (2009), Bayesian disease mapping, CRC press.
6. Mollie, A. (1994). Bayesian Mapping of Disease. In: Gilks, W., Richardson, S., and Spiegelhalter, D. J. (eds.). Markov Chain Monte Carlo in Practice. London: Chapman and Hall.
7. Montesanto A., Passarino G., Senatore A., Carotenuto L., De Benedictis G, (2008), Spatial Analysis and Surname Analysis: complementary Tools for Shedding Light on Human Longevity Patterns, Ann. Hum. Genet. 72: 253-260.
8. Miglio R., Marino M., Rettaroli R., Samoggia A., Spatial analysis of longevity in a Northern Italian region, paper presented at the PAA Annual Conference, Detroit, MI, 28/4-2/5 2009
9. Poulain M., Pes G., Grasland C., Carru C., Ferrucci L., Baggio G., Franceschi C., Deiana L. (2004), Identification of a geographic area characterized by estreme longevity in the Sardinia Island: the AKEA study, Exp. Gereontol. (39): 1423-1429.
10. Robine, J. M., Caselli, G.,(2005) An unprecented increase in the number of centenarians, Genus, LXI: 57-82.
11. Robine, J. M. , Paccaud, F. (2005) Nonagenarians and centenarians in Switzerland, 1860-2001: a demographic analysis. J. Epidemiol. Comm. H. 59:31-37.
12. Thatcher, R. A. (2001) The demography of Centenarians in England and Wales. Population-An English Selection, INED. 13(1):1-139.