This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

LONGPOP

Methodologies and Data mining techniques for the analysis of Big Data based on Longitudinal Population and Epidemiological Registers

Compilation of GIS layers on Italian Demography

Deliverable n. 2.3

Introduction

Italy has always been a country characterized by striking territorial cultural and socio-economic disparities and a very pronounced geographical variation from a demographic point of view. This heterogeneity makes the country a very appropriate case study for applying GIS and spatial analysis techniques to the demographic behaviours.

An increasing amount of research efforts have been devoted in the last years to refining the statistical methods for analysing mortality in small geographical areas (Divino et al. 2009). These analyses not only allow explore the distribution of mortality in detail, but also help analyse the relationship between mortality and some contextual variables that could, at least in part, explain the health status of a population, such as the socioeconomic or environmental variables.

In this regard, Italy is a particularly interesting case. The country is indeed well known for being amongst the ones with the highest life expectancy at birth, but its remarkable geographical mortality variation has never been properly explored at the municipal level.

To our knowledge, previous researches have been mainly conducted at the regional (NUTS 2) or provincial level (NUTS 3) [Egidi et al. 2005]. Furthermore, the few studies carried out at the municipal level were limited to a single region [Biggeri et al. 2006].

In a first phase, due to the peculiar and unique features of mortality transition in the island, our analyses were focused exclusively on Sardinia. The extreme poverty and very low literacy of the population, as well as its delayed process of industrialization, make the health transition process in the island particularly interesting and intriguing. Since Italian unification, in fact, Sardinia has been one of the most economically backward regions in the Peninsula. The main agricultural activity suffered from an extreme landholding fragmentation, due to the pursuit of exasperated egalitarianism in hereditary. The other main economic activity was shepherding, mainly practised in the mountainous interior zones. The unique economic activity ascribable to the industrial sector was mining. In particular, coal extraction in the Sulcis, lead and zinc mining in the Iglesiente (both in south-west Sardinia), in addition to scattered quarrying elsewhere gave employment up to the 9% of the population at the peak of mining employment in 1939 during the fascist regime.

In the aftermath of World War II, the Italian Government decided to launch an ambitious program of public incentives to promote the industrialization of the Southern of the country with the creation of the Fund for Southern Italy, an institution aimed at financing the improvement of infrastructures and sustaining industrial activity by furnishing cheap credit. Sardinia was obviously identified as an objective for the program. This led in 1962 to the so called Plan for the Rebirth of Sardinia which in contrast to the Cassa's policy of favouring the modernization of the agrarian activity and infrastructural investments, was mainly focused in promoting the localization of heavy industry in the island.

Originally, 18 sites (after reduced to 15) were identified as appropriate zones for spending the huge amount of public and private investments programmed for the project (around 1800-2000 billions Italian lire).

One of the largest European petrochemical complex was built in Porto Torres (a coastal village in the North western part of the region) in the first half of the 1960s. Contemporary, an oil refinery was built in Sarroch (near the capital town of the island, Cagliari). Other smaller scale industrial activities were constructed elsewhere. Finally, in 1969, another large chemical plant was localized in Ottana (between the area of Macomer and the zone of Nuoro). Given the poor environmental awareness of the epoch, the industrial plants financed by the Rebirth Plan were generally highly pollutant.

Indeed, two of these industrial sites, Porto Torres and Portoscuso, have been recently classified according to the Italian law as "Siti di Interesse Nazionale", that is, territories identified as contaminated in relation to both the quantity and dangerousness of the present pollutants and for the impact they can have on the surrounding environment, in terms of

health and ecological risks. In addition to industrial sites also the military base of La Maddalena (an island in the North East of Sardinia) is considered at high environmental risk.

Results

The aim of our research in the LONGPOP project has been that of investigating the long-term consequences on mortality of the Sardinian plan of rebirth, by the means of a spatial analysis carried out at the municipal level both on the eve of the industrialization process (from 1951 to 1961) and after 20, 30, 40 and 50 years from the beginning of the plan.

We created a longitudinal database for the long-term reconstruction of Sardinian mortality geography, at the municipal level, for the period 1992-2015, then we have extended the analysis back in time. From these data, we first created GIS layers of mortality using spatial data analysis techniques, and thereafter we analysed the relationship of mortality with contextual socioeconomic variables.

The procedure we followed could be briefly described as follows. We calculated standardized mortality ratios (SMRs) as the ratio between observed and expected cases.

To calculate the number of expected cases, we multiplied the Sardinian age-sex specific mortality rates for by the municipal population for each sex and age group. We also calculated the 95% confidence intervals of the SMRs. Smoothed municipal relative risks (RRs) with their corresponding 95% credibility intervals and posterior probabilities (PRPs), that smoothed RR was greater than one, were calculated using the conditional autoregressive model proposed by Besag, York and Mollié. This model adjusts a spatial Poisson model with two types of random effects, an unstructured effect, which accounts for unstructured heterogeneity, and a structured effect, the spatial term, which considers the contiguity between areas.

Regarding PRPs, following the criterion proposed by Richardson, we consider as "municipalities with high mortality risk" those with a value equal or greater than 0.8, and as "municipalities with low mortality risk" those with a value equal or less than to 0.2. The tool used for Bayesian inference of the subsequent marginal distributions was Integrated Nested Laplace Approximations (INLA). For this, we used the R-INLA library available in the R statistical package. In order to create the maps, we used version 10.5 of the ArcMap software package.

See Fig. 1 and below Fig. 2 for examples referring to Smoothed municipal relative risks (RRs) for the female population in the first and last period analysed.
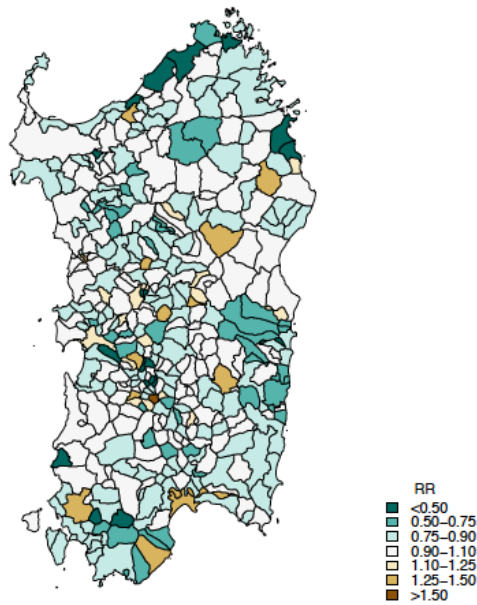
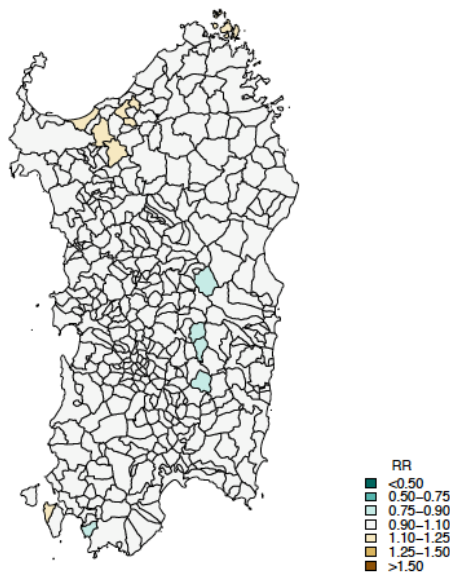**Fig. 1 Smoothed municipal relative risks for the Female Population, Sardinia (1951-1961).**

Fig.2 Smoothed municipal relative risks for the Female Population, Sardinia (2010-2015).

In general, the results (not reported here) of spatial analysis did not indicate any prevailing geographical pattern in RRs for mortality throughout the periods of the study for women or men. For both sexes, we observe, in the first period analysed, higher SMRs in the urban areas and mining sites. In the recent periods, the largest mortality rates were recorded in the industrial area of Porto Torres. When PRP are considered, in the last period a marked geographical pattern arises, with numerous high-risk municipalities in the North-Eastern areas of the island (Sassari and Olbia) together with a low-risk zone in the central area of the region (Province of Nuoro).

During the last months of the LONGPOP project, we have extended this analysis by creating the Italian Municipal Database (IMD). The work is still in progress but we have already at our disposal the results from 1992. This database is based on the census data provided by the Italian Institute of Statistics at the municipal level, extending to the whole country the previous analysis focussed on Sardinia. Corresponding GIS mortality layers have been created using the same approach described above, in order to identify spatial mortality patterns. We have also analysed the relationship between mortality and contextual factors, such as unemployment. Lastly the Sassari's team has expanded the IMD to include mortality data and contextual factors from the 1951, 1961, 1971, 1981, 1991 and 2001 Census. See Fig. 3 and 4 for examples referring to smoothed standardized mortality ratios in Italy for the male and the female population in 2010-2016.
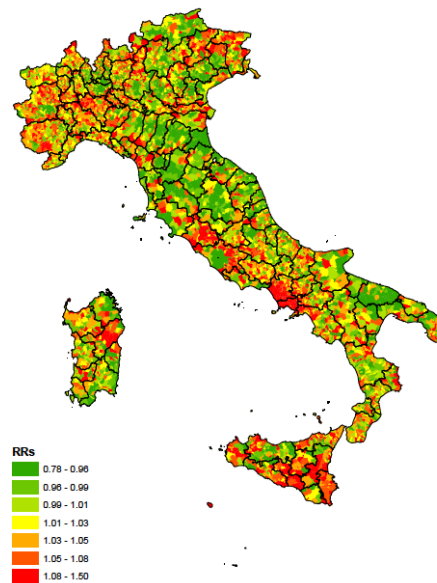

Fig.3 Smoothed standardized mortality ratios for the MALE Population, Italy (2010-2016)
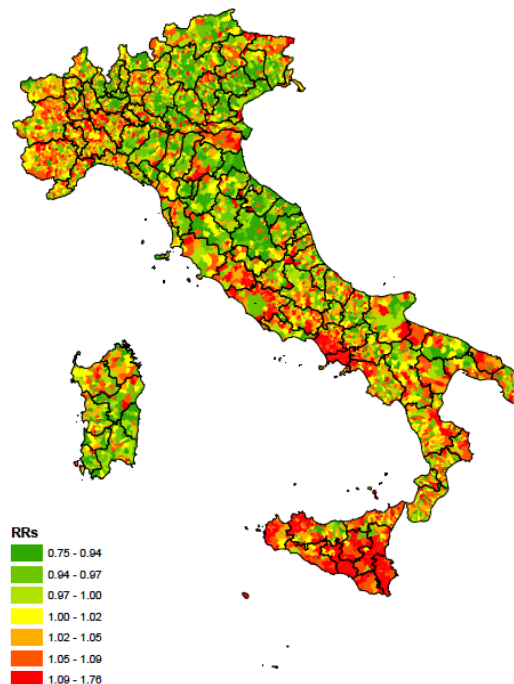
Fig.4 Smoothed standardized mortality ratios for the FEMALE Population, Italy (2010-2016)

The first map shows municipalities with higher mortality in the southwest of the Peninsula, specifically in the regions of Campania (Provinces of Caserta and Naples) and Latium (Province of Viterbo). The province of Nuoro in Sardinia also stands out, as well as the southeast part of Sicily. The spatial mortality pattern of women is very similar to that identified for men. The Southern regions of Campania and Sicily remain the areas characterized by the highest mortality, although with respect to the male counterpart, in Sicily there are more provinces with elevated mortality. Also noteworthy is the situation of the Apulia region, located in the south of the Peninsula. All these regions are characterized for both men and women by a statistically significant overall mortality.

Moreover, shapefiles are being created with the territorial demarcations of the municipalities from 1951 onwards, which will allow us to create GIS mortality layers for previous years and to study the geographical evolution of mortality in relation to several factors.

As regards to the analysis mobility and nuptiality, we have created preliminary descriptive GIS layers for the whole Italian country, at the municipal level for the period 2010-2016. See Fig. 5 and 6 for examples with reference to both migration and marriages.
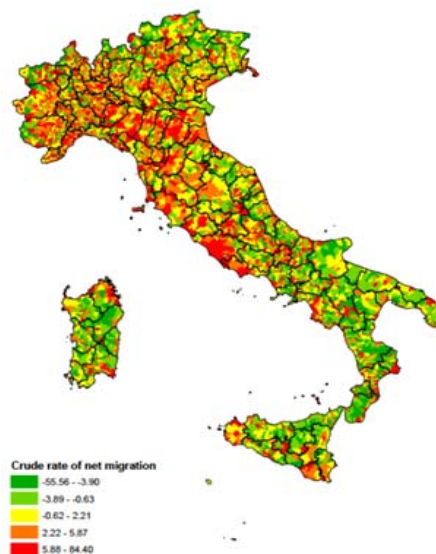

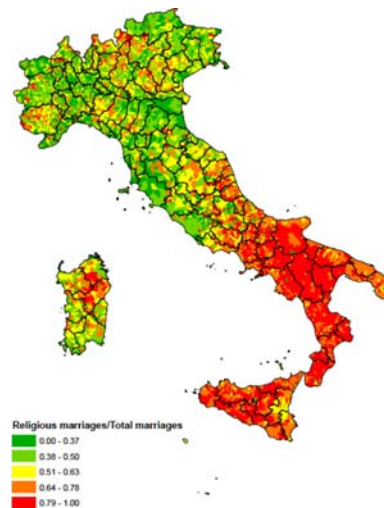**Fig. 5 Crude rate of net migration, Italy (2010-2016)**

Fig. 6 Religious Marriages, Italy (2010-2016)

Given the large economic divide between Northern and Southern Italy, it is not surprising that in vast part of Southern Italy, net migration is negative while the opposite is found for the municipalities in North-Western Italy and in a lesser extent in Latium, where Rome is the main attraction basin (see Bonifazi 2015 for a discussion).

This last figure shows a clear North-South divide and a much higher diffusion of religious marriages in the South, which is probably one of the most evident and emblematic expressions of the profound cultural divisions still existing in the country. Sardinia is a very interesting case, having become in the last decades a region much more similar to those of the north and the centre of the country from the demographic point of view, particularly in terms of marital and reproductive behaviours.

Lastly, as regards to the analysis of nuptiality and mobility, we are currently exploiting an exhaustive database of marriages in the Friuli region (province of Udine). This database (kindly made available by A. Fornasin) covers over 80,000 individual acts relating to marriages that took place between 1872 and 1900 in the 127 municipalities that were then present in the analysed territory. In addition to recalling the dominant features of the regional marriage model, at the spatial level, greater attention will be paid to the differential analysis according to the rank of marriage (first or subsequent).

The Sassari team is in the process of ensuring that an IT platform is available on the department's web page where the Gis layers and the data will be soon accessible.

References

Besag J, York J, Mollié A. Bayesian image restoration with two applications in spatial statistcs. Ann Inst Stat Math. 19 91;43:1-59.

Besag J. Spatial interaction and the statistical analysis of lattice systems. J R Stat Soc Series B Stat Methodol. 1974;36(2):192-225.

Biggeri A, Lagazio C, Catelan D, Pirastu R, Casson F, Terracini B. Ambiente e salute nelle aree a rischio della Sardegna. Epidemiol Prev 2006; 30(1) Suppl 1: 1-96.

Bonifazi, C. 2015. Le migrazioni tra Sud e CentroNord: persistenze e novità. In La nuova migrazione italiana. Cause, mete e figure sociali. Ed.: Gjergji, I., Venezia, Edizioni Ca' Foscari, 57–69.

Clayton D, Bernardinelli L, Montomoli C. Spatial correlation in ecological analysis. Int J Epidemiol. 1993;22(6):1193-202.

Divino F, Edigi V, Salvatore MA. Geographical mortality patterns in Italy: A Bayesian analysis. Dem Res. 2009; 20(18):435-466.

Egidi V, Spizzichino D, Frova L, Pappagallo M, Di Fraia G. Sopravvivenza e mortalità nelle province italiane. Rapporto osservasalute 2005,; 2005:43-67.

Richardson S, Thomson A, Best N, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies, Environ. Health Perspect. 2004;112:1016-1025.

R Development Core Team: R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2005.

Rue H, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). J R Stat Soc Series B Stat Methodol. 2009;71(2):319-392.Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc Series B Stat Methodol. 2002;64:583-616.The R-INLA project. http://www.r-inla.org/. Accessed 20 June 2019.

Best N, Richardson S, Thomson A. A comparison of Bayesian spatial models for disease mapping. Stat Met Med Res. 2005;14:35-9.