International Journal of Mathematical Sciences in Medicine (2013) 6-16

Cancer mortality in 180 countries of the world investigated through spatial autocorrelation models. Is there a clustered pattern of cancer?

Sifaki-Pistolla Dimitra[1], Pistolla Georgia [2], Anastas Philalithis[3]

[1] Geographer, Master in Public Health and Epidemiology in Department of Social Medicine, Faculty of Medicine, University of Crete Voutes, 2208, Greece

[2] Mathematician, PhD in Department of Social Medicine, Faculty of Medicine, University of Crete, Voutes, 2208, Greece

[3] Professor in Social Medicine, Department of Medicine, University of Crete, Voutes, 2208,

Greece

Abstract: Cancer is a leading cause of death worldwide. Purpose of the present study is to investigate the spatial autocorrelation of cancer deaths in 180 countries worldwide. It aims at answering whether neighboring countries tend to have similar cancer deaths or not and how these 180 countries could be clustered according to their spatial autocorrelation.

Methods: Data of total cancer deaths and cancer deaths for both genders per 100,000 people were selected from World Health Organization database. Cancer deaths were adjusted to age and selected for 180 countries worldwide. The whole analysis was carried out in the GIs Arc map and spatial statistics. Moran's I was applied in order to detect the spatial autocorrelation.

Results: Positive spatial autocorrelation was proved as far as cancer mortality is concerned, worldwide. Total Moran's I= 0.10, men's Moran's I= 0.13 and women's Moran's I = 0.05 (p=0.01). There is a spatial clustering of the 180 countries which measures their correlation and indicates that neighbouring countries tend to have similar values of cancer mortality.
Conclusions: This study imports new methods of analysis of spatial data in Epidemiology. Therefore, it highlights the possible applications of these methods as well as their suitability in capturing epidemiological issues.

Key words: cancer mortality, spatial autocorrelation, Global Moran's I, spatial clusters, spatial epidemiology

Received August 15, 2012 IFNA Publisher, Inc

1 INTRODUCTION

Cancer is a leading cause of death worldwide. It is estimated that 7.6 million deaths (around 13% of all deaths) occurred in 2008. Lung, stomach, liver, colon and breast cancer are the most frequent cancer deaths each year. At the same time the most frequent types of cancer differ between men and women, spatial regions, time periods and age (Globocan, 2010). WHO estimates that 84 million people may die from cancer from 2005 to 2015 with no intervention, while 30-40% can be treated effectively. This indicated that proper prevention or treatment measures could help in reducing the incidence of cancer (WHO, 2011; Christopher & Lopez, 1997). Risk factors play an important role for cancer incidence. About 30% of cancer deaths are due to the five leading behavioral and dietary risks: high body mass index, low fruit and vegetable intake, lack of physical activity, tobacco use and alcohol use (WHO, 2011).

Reliable information on causes of death is essential to the development of national and international health policies for prevention and control of disease and injury. Non-communicable diseases are, however, already major public health challenges in all regions (Christopher & Lopez, 1997). Cancer researchers have a long tradition of presenting data in a spatial representation, or maps in order to communicate population-based cancer statistics. Starting from J. Snow, health data used to be presented in printed atlases, as a try to apply a kind of spatial analysis for epidemiological issues (Bell, 2006; Boscoe et al., 2004; Buntix et al., 2003; Jacquez, 2004).

Nowadays, these applications are more developed and able to offer a reliable, sophisticated analysis over several phenomena. They can produce distribution maps for monitoring of a disease, spatial prediction for the dynamic of a disease, spatial statistics, spatial autocorrelation maps and maps of risk areas (Gloud & Wallace, 1994). Cancer mortality spread could be investigated by applying modern spatial analysis following several steps (Getis & Ord, 1992; Anselin, 1995; Anselin, 1993). The first step could be searching the spatial autocorrelation of the disease worldwide. Then, correlation of the risk factors should be explored, estimation of the risk areas and areas for emergence intervention should be applied and finally, prediction of the spatial variance even where there no data available (Stein et al., 2002; Kamran et al., 2012; Cliff & Ord, 1981).

Purpose of the present study is to investigate the spatial autocorrelation of cancer deaths in 180 countries worldwide. Basic research question is whether neighboring countries tend to have similar cancer deaths or not and how these 180 countries could be clustered according to their spatial autocorrelation. Final goal of this study is to import new methods of analysis of spatial data in Epidemiology. Therefore, it is integral to highlight the possible applications of these methods as well as their suitability in detecting or analyzing epidemiological issues or dealing with Public health strategies of intervention and management.

2 Material and Methods

Data of total cancer deaths and cancer deaths for both sexes per 100,000 people were selected from World Health Organization database.[15] Cancer deaths were adjusted to age and selected for 194 countries worldwide. Countries with no data were excluded, concluding to 180 countries. Geo-data of countries' coastline were used in order to analyse the cancer deaths spatially and they were geo-referenced in the EGSA 87 (Greek National Georeference

datum 87 with projection: Greek Grid). Both sets of data were imported and analysed in GIs Arc map 9.2.

Three spatial distribution maps were created in order to present the spatial variance of cancer deaths in these 180 countries for total cancer deaths as well as for men and women respectively. In order to achieve this, six clusters of number of cancer deaths were exported by the spatial method of natural breaks. This method creates a classification of the data. It seeks to partition data into classes based on natural groups in the data distribution. Natural breaks occur in the histogram at the low points of valleys. Breaks are assigned in the order of the size of the valleys, with the largest valley being assigned the first natural break (Jenks, 1967).

Then, spatial statistics were applied in order to examine the spatial autocorrelation the total cancer deaths, cancer deaths among men and among women, worldwide. Moran's I with filters of Global Moran's index, Inverse distance and Euclidean distance was applied. Techniques which test the spatial dependence among values of different spatial units, consist a whole separate category of analyzing geo-epidemiological data. They focus on spatial autocorrelation and they may be local or global methods. Such a method is the one applied in this study: Global Moran's I.

Therefore, it is a measure of spatial autocorrelation developed by Patrick A.P. Moran. Spatial autocorrelation is characterized by a correlation in a signal among nearby locations in space. Spatial autocorrelation is more complex than one-dimensional autocorrelation because spatial correlation is multi-dimensional (i.e. 2 or 3 dimensions of space) and multi-directional (Getis & Ord, 1992; Anselin, 1995; Anselin,1993; Jenks, 1976; Moran, 1950)

Moran's I is defined as:

$$I= N/ \Sigma_i \Sigma_j w_{ij} * \Sigma_i \Sigma_j w_{ij} (\overline{X}_i-X) (\overline{X}_j-X)/ \overline{\Sigma}_i (Xi-X)^2$$

where N is the number of spatial units indexed by i and j, X is the variable of interest, $\overline{X}$ is the mean of X and w_{ij} is an element of a matrix of spatial weights.

The expected value of Moran's I under hypothesis of no spatial autocorrelation is:

$$E (I) = -1/N-1$$

Its variance is equal to:

$$Var (I) = NS_4-S_3S_5/ (N-1)(N-2)(N-3)(\Sigma_i \Sigma_j w_{ij})^2$$

Where:

$$S_1=1/2 (\underset{i\ j}{\Sigma\Sigma} (w_{ij} + w_{ji})^2)$$

$$S_2= \Sigma_i (\Sigma_j w_{ij} + \Sigma_j w_{ji})^{\ 2} / 1$$

$$S_3= N^{-1} \Sigma_i (x_i-\overline{x})^4 / (N^{-1} \Sigma_i(xi-\overline{x})^2)^2$$

$S_4= (N^2-3N+3)\ S_1-NS_2+3(\Sigma_i\Sigma_j\ w_{ij})^2 / 1$
$S_5= S_1-2NS_1+6\ (\Sigma_i\Sigma_j\ w_{ij})^2 / 1$

Negative values indicate negative spatial autocorrelation and positive values indicate positive autocorrelation. Values range from −1 (indicating perfect dispersion) to +1 (perfect correlation). Finally, a zero value indicates a random spatial pattern. For statistical hypothesis testing, Moran's I values can be transformed to z-scores in which values greater than 1.96 or smaller than −1.96 indicate spatial utocorrelation that is significant at the 5% level (Getis & Ord, 1992; Anselin, 1995; Anselin,1993).

In the present study "w" was measured by the Inverse distance weighting (IDW) is a method for multivariate interpolation, a process of assigning values to unknown points by using values from usually scattered set of known points. Therefore, the value at the unknown point is a weighted sum of the values of N known points.
A general form of finding an interpolated value u at a given point x based on samples $u_i = u(x_i)$ for i = 0,1,...,N using IDW is an interpolating function:

$$u(x)= \sum_{i=0} [\ w_i(x)\ u_i / \sum_{j=0} w_j\ (x)]$$

Where:

$$w_i(x)= 1 / d(x,x_i)^p$$

is a simple IDW weighting function, as defined by Shepard D., x denotes an interpolated (arbitrary) point, x_i is an interpolating (known) point, d is a given distance (metric operator) from the known point x_i to the unknown point x, N is the total number of known points used in interpolation and p is a positive real number, called the power parameter.

Here weight decreases as distance increases from the interpolated points. Greater values of p assign greater influence to values closest to the interpolated point (Shepard, 1968).

Finally, the results of this test are presented in the next section as three maps of spatial autocorrelation clustering, as well as the produced Moran scatter plot for the three groups of data (total cancer deaths, cancer deaths among men and women).

3 RESULTS

In figure 1, box plots are presented, whereas in figure 2 spatial variation of cancer mortality is distributed. It is obvious that there is spatial heterogeneity among these countries. Values range from 94 to 426 deaths per 100,000 people as far as cancer deaths for both sexes are concerned. Cancer deaths among men vary from 39 to 260 deaths per 100,000 people and 40 to 191 deaths per 100,000 people among women. Furthermore, a mean value of 223.43 deaths per 100,000 people (St.dev = 59.095) is estimated for cancer deaths for both sexes [mean $_{men}$ = 127.74 (St.dev= 44.426) and mean $_{women}$ = 96.01 (St.dev= 21.705)]. Men present higher rates of cancer deaths on comparison to women (figure 1). Number of cancer deaths also differs from one country to another.

Higher values of total deaths per 100,000 people are presented at Armenia (N=363), Croatia (N=340), Hungary (N=389), Latvia (N=342), Lithuania (N=330), Mongolia (N=426), Poland (N=350), Serbia (N=340), South Africa (N=331) and Uruguay (N=335). Countries that follow with their high levels of cancer deaths are: Albania, China, Belgium, Cuba, Denmark, Germany, France, Ireland and other European countries. On the other hand, countries with lower rates are Gabon, Cook Islands, Kiribati, Kuwait, Maldives, Namibia, Samoa, Syrian Arab Republic, Unites Arab Emirates and other countries of Africa (starting from 94 to 124 deaths per 100,000 people). North America, Australia and Brazil are at medium levels of cancer mortality (figure 2).

Finally, differences between genders are obvious even at the same country. For instance, Greece has a rate of 252 cancer deaths per 100,000 people (165 deaths among men and 87 deaths among women). Russia presents the highest difference between genders. There are 194 cancer deaths among men and 89 deaths among women (figure 2).

Figure 1: Box plots of number of cancer deaths per 100,000 people in 180 countries worldwide, in 2009.

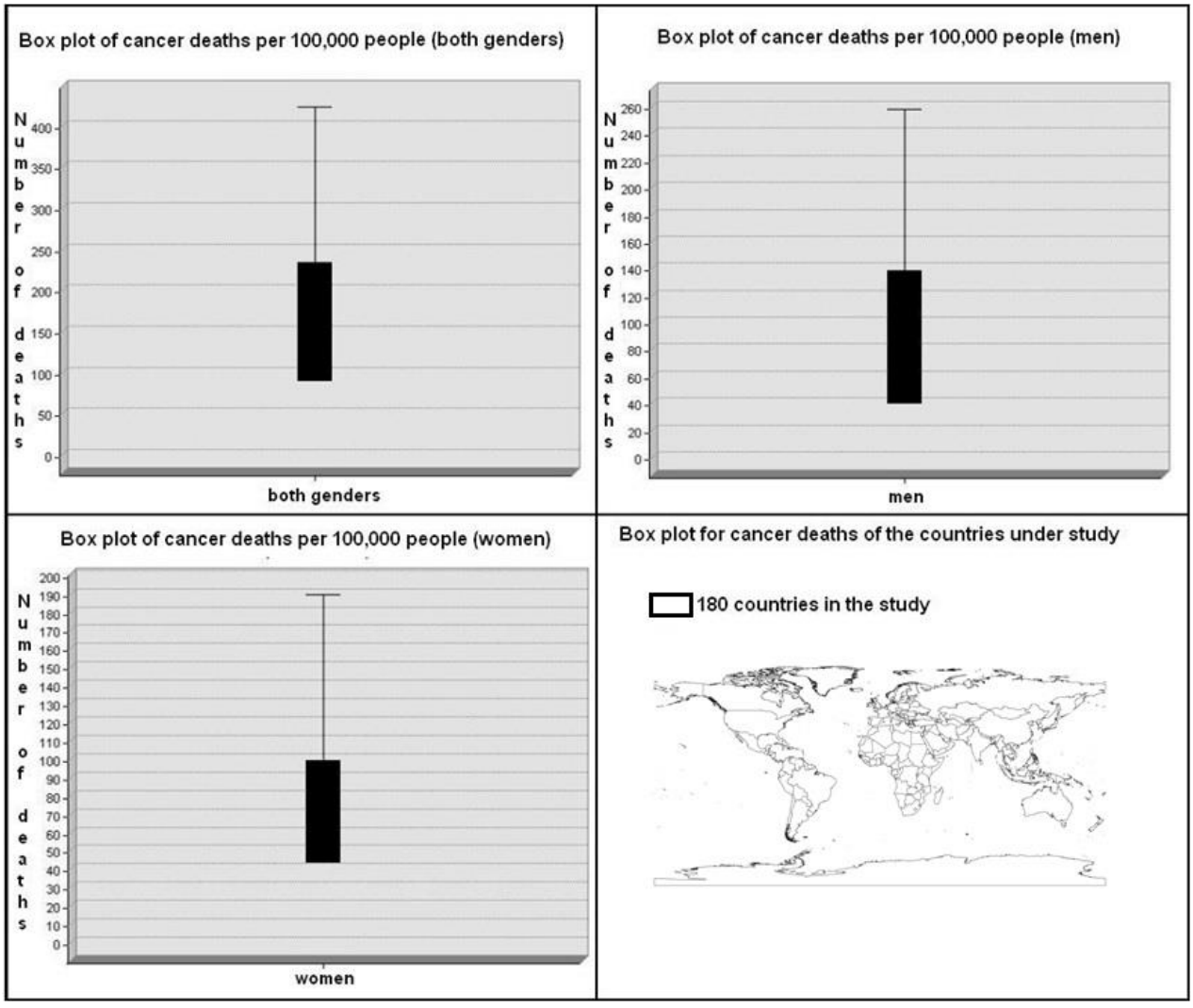
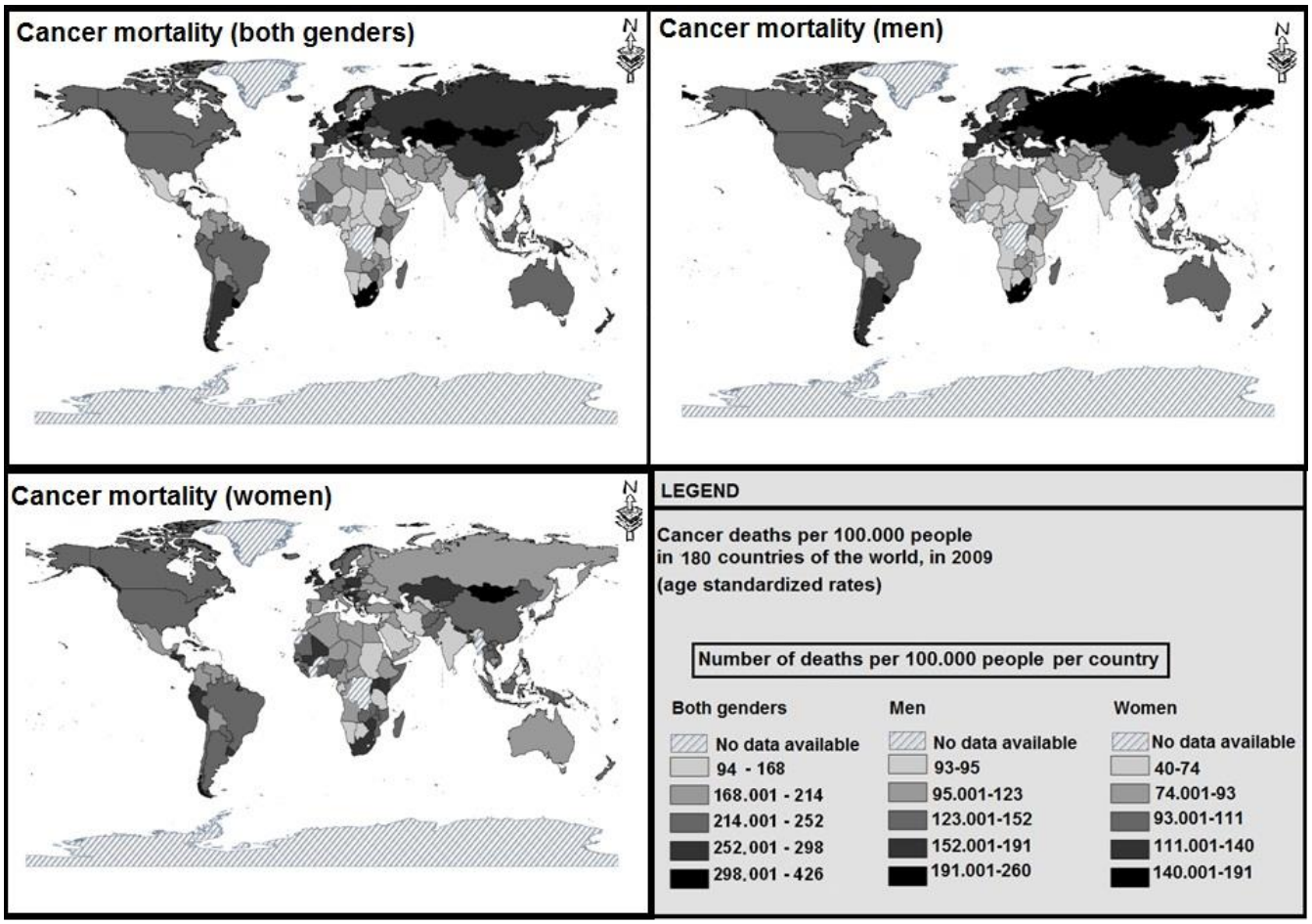


Figure 2: Spatial distribution of Cancer deaths per 100,000 people in 180 countries worldwide, in 2009.



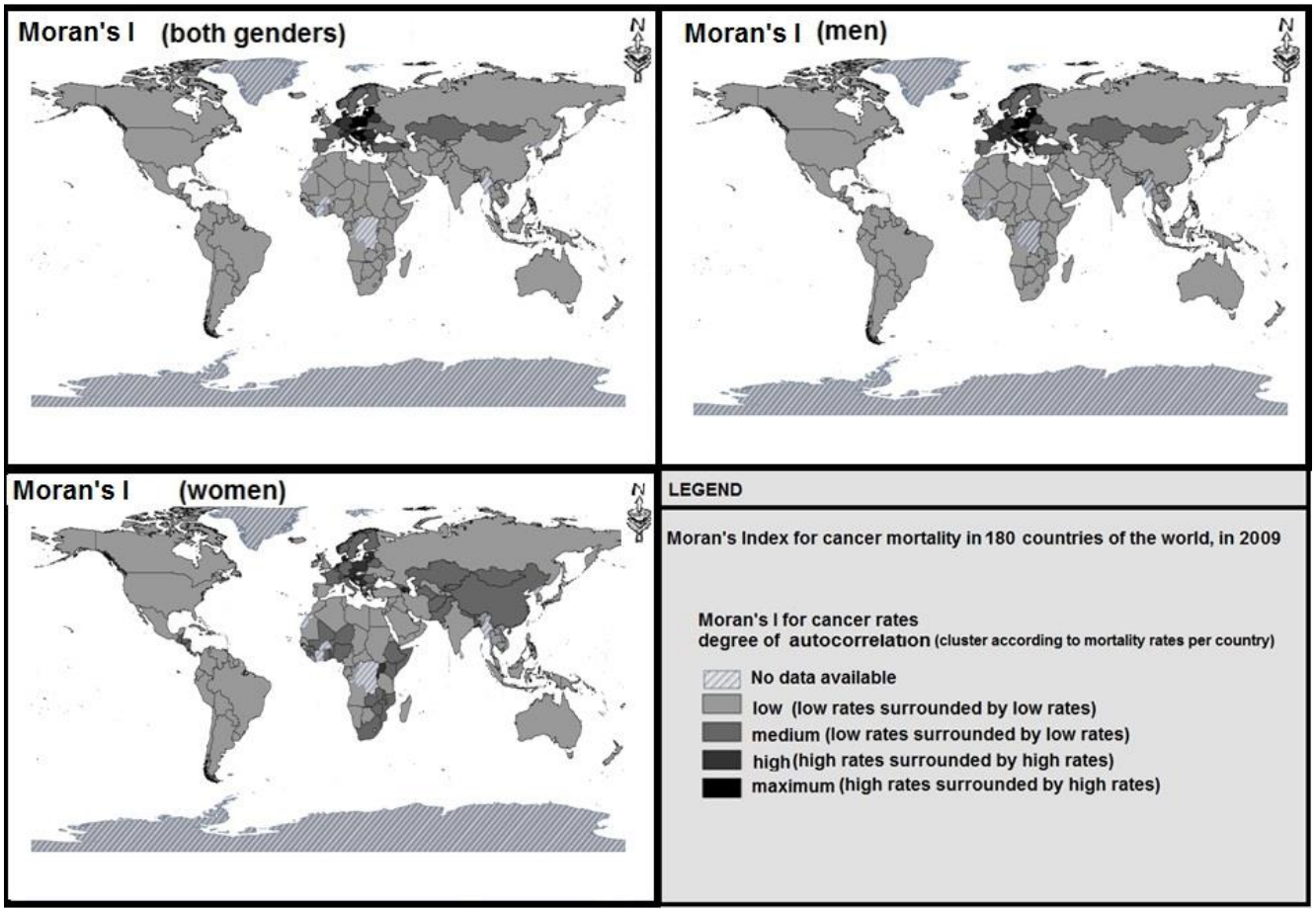
In figure 3, spatial autocorrelation is demonstrated. Spatial clusters are indicating the positive spatial autocorrelation and they are as of low–low (low rates in the areas surrounded by areas of low rates), low-high (low rates surrounded by high rates), high-low (high rates surrounded by low rates) and high-high (high rates surrounded by high rates). In the present study there are only high-high and low-low clusters. In order to make a grouping of these clusters in figure 3, we re-named them as of low, medium, high and maximum degree of autocorrelation.

Moran's I for total number of cancer deaths per 100,000 people was equal to 0.10 (standard deviation =6.35, p=0.01). Moreover, variance was estimated equal to 0.0002, critical value = 2.58 and expected index = -0.004. The pattern of the cancer deaths' distribution was proved to be clustered.

Similar results were exported for cancer deaths by gender. Moran's I for men's cancer deaths per 100,000 people was equal to 0.13 (Stdev. = 9.28), variance = 0.0002, critical value= 2.58 and expected index = -0.004. As far as women's cancer deaths are concerned, Moran's I = 0.05 (Stdev. = 2.87), variance = 0.0002, critical value= 2.58 and expected index = -0.004. There is less than 1% likelihood that these clustered patterns (for total number of cancer deaths, men and women) could be the result of a random chance. Consequently, positive spatial autocorrelation was proved with men and cancer deaths of both sexes presenting a higher positive spatial autocorrelation. Positive spatial autocorrelation has all similar values appearing together. The existence of positive spatial autocorrelation in these

data refers to a map pattern where geographic features of similar value tend to cluster on a map. Neighboring regions tend to have similar cancer mortality ratios than countries which are far away. For instance, America, Africa, Australia, Russia and some other regions of Asia are grouped together to the lowest positive correlation cluster and present similar behavior as far as total cancer mortality is concerned. North Europe, Turkey and Spain are grouped together at the medium level of positive autocorrelation whereas central and southern Europe is at the high level, apart from Latvia, Lithuania, Poland and Serbia which are clustered at the maximum level of spatial autocorrelation.

Figure 3: Spatial autocorrelation – Moran's I applied on cancer deaths per 100,000 people in 180 countries worldwide, in 2009.

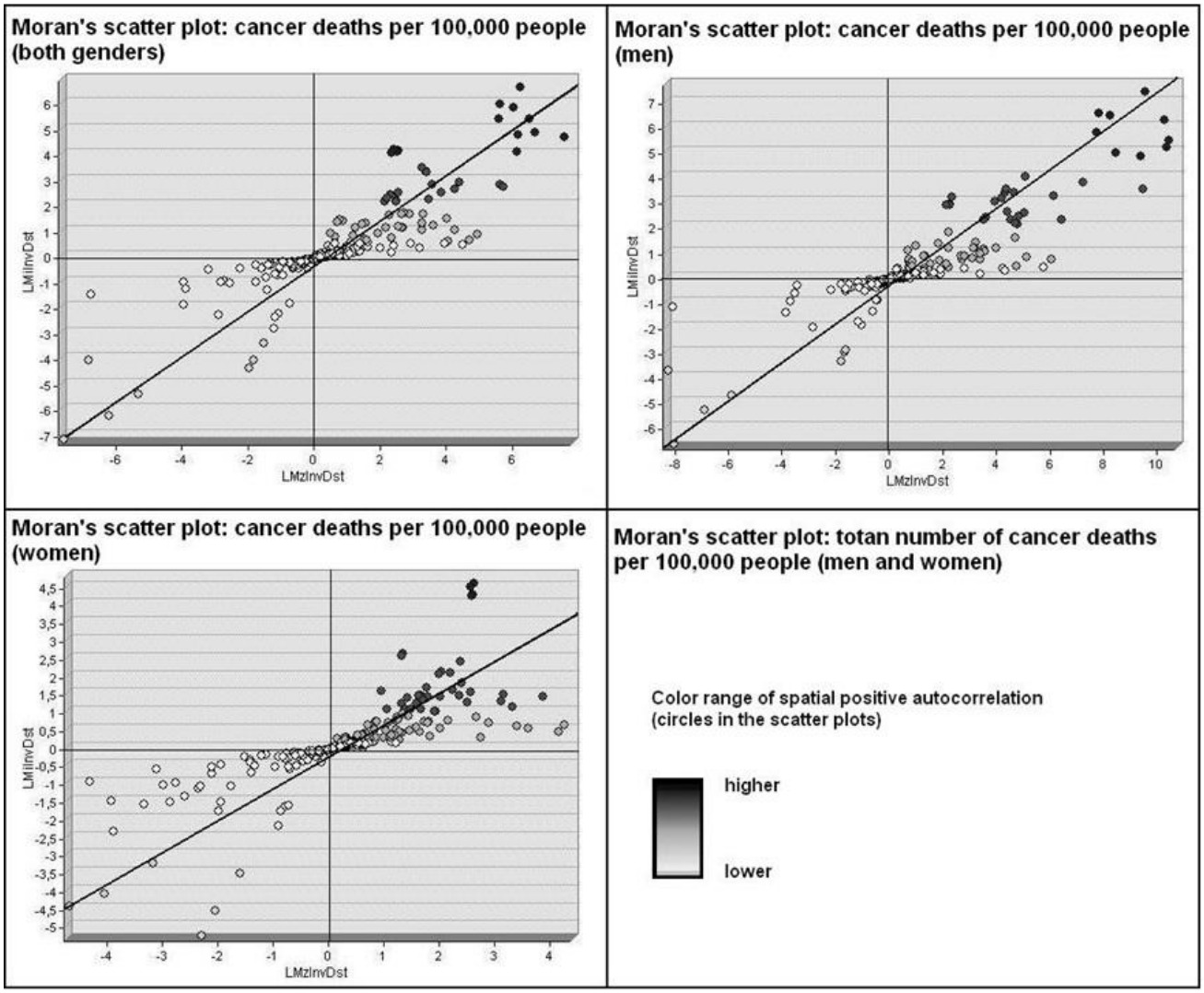


The Moran scatter plot (figure 4) provides a tool for visual exploration of spatial autocorrelation which was presented in figure 3. It is actually the spatial lag of the variable on the vertical axis and the original variable on the horizontal axis. The spatial lag refers to the values of a location's neighbors.

The standardized variables (not the raw data) are visualized in figure 4. Standardization is denoted by a (Z), for *z*-score, after the variable name. Different colors are used in order to present the degree of spatial autocorrelation creating clusters. As it was mentioned before, there are only low-low and high – high clusters since there are values only in the upper right and lower left quadrant of the scatter plot. Different colors are used in order to identify this,

as well as their degree of positive spatial autocorrelation. Clusters characterized as maximum [high-high] (upper right quadrant), low [low-low] clusters (lower left quadrant), medium [low-low] (lower left quadrant) and high [high-high] clusters (upper right quadrant).

Figure 4: Moran's scatter plot for cancer deaths per 100,000 people in 180 countries worldwide, in 2009.



4 DISCUSSION

Awareness about the causes of cancer, and interventions to prevent or deal with the disease is extensive. Cancer can be reduced and controlled by implementing evidence-based strategies for cancer prevention, based on reliable studies over cancer variance worldwide and its spatial spread.

Spatial analysis could be a tool for achieving the above and dealing with several epidemiological problems. The spatial parameter exists in every phenomenon on earth and should be tested, especially in the field of Epidemiology and Public Health. A reliable way to study a phenomenon as close to its natural environment as it is possible is spatial environments and softwares, such as GIs Arc map. There are plenty of methods and

strategies for exploring diseases and mortality. In the preset study, we start with basic spatial applications (such as distribution maps) and we go on with sophisticated spatial statistics, related to spatial autocorrelation (Gould, 1994).

The use of spatially referenced data in cancer studies is gaining in prominence, fuelled by the development and availability of spatial analytic tools and the broadening recognition of the linkages between geography and health (Gloud, 1994; Miller & Wentz, 2993; Lacquez, 2004). Spatial analysis is valuable for data exploration, identification of geographic patterns, generation of new hypotheses, and providing supporting evidence about existing hypotheses. A geographic perspective will be increasingly relevant as GIS software, spatial analytic methods, and the availability and quality of geographically referenced data continues to improve (Miller & Wentz, 2993; Lacquez, 2004). These improvements are applied on various diseases and epidemiological issues helping in reliable research and studies.

Cancer mortality is an epidemic in modern ages which is constantly studied. Its spatial status is extremely important and not completely known, yet. The cancer mortality's spatial inequalities are immense even in the same continent or country. It is utterly characterized by a local spatial pattern (Peto et al., 1992; Pistolla et al., 2010).

Positive spatial autocorrelation was proved as far as cancer mortality is concerned, worldwide. There is a spatial clustering of the 180 countries which measures their correlation and indicates that neighbouring countries have similar values. Consequently, cancer mortality is not a random phenomenon. This is a useful piece of information in order to give an incentive for further research, apply prevention or management policies in terms of Public Health and understand the behaviour of the under study phenomenon.

Spatial correlation of basic risk factors (such as cholesterol, alcohol, tobacco, mean blood glucose, blood pressure, Body Mass Index, UV radiation exposure, median age per country, population aged over 60, outdoor and indoor pollution exposure and physical inactivity) or other factors (total expenditure on health/ per capita, expenditure for medical equipment, cancer registries per country), estimation of risk areas and spatial regression models are under study right now and would be presented in our next paper.

5 CONCLUSIONS

We suggest extensive spatial studies of several phenomena such as diseases and causes of mortality. Further study should be carried out on cancer mortality and its correlation with various risk factors, concluding to the detection of risk areas worldwide or at a local level. Suggested methods for the next step are: non-linear spatial regression models, multicritiria analysis and hot spot trends.

6. REFERENCES

[1] A. Cliff and J.K. Ord, Spatial Processes: Models and Applications, (1981) London, Pion.

[2] A.F. Stein, van der Meer and B. Gorte, Spatial Statistics for remote sensing. Environmental and ecological statistics, 83 (2002) 17-33.

[3] A. Getis and J.K. Ord, The analysis of spatial association by use of distance statistics, Geographical analysis, 24 (1992) 289-206.

[4] D. Shepard, A two-dimensional interpolation function for irregularly-spaced data, Proceedings of the 1968 ACM National Conference, (1968) 517–524.

[5] F. Buntinx, H. Geys, D. Lousbergh, G. Broeders, E. Cloes, D.L. Dhollander, Op De Beeck, J. Vanden Brande, A. Van Waes, G. Molenberghs, Geographical differences in cancer incidence in the Belgian province of Limburg, European Journal of Cancer, 39 (2003) 2060:2065.

[6] F.P. Boscoe, M.H. Ward and P. Reynolds, Current practices in spatial analysis of cancer data: data characteristics and data sources for geographic studies of cancer, International Journal of Health Geographics, 3 (2004) 28.

[7] G.F. Jenks, The Data Model Concept in Statistical Mapping, International Yearbook of Cartography, 7 (1967) 186–190.

[8] Globocan 2008, IARC, 2010, available at: http://globocan.iarc.fr/factsheets/populations/factsheet.asp?uno=900, Accessed in: 28/01/2012

[9] G.M. Jacquez, Current practices in spatial analysis of cancer data: flies in the ointment, or, the limitations of spatial analysis, International Journal of Health Geographics, 3 (2004) 22.

[10] G. Pistolla, P. Prastakos, M. Vassilaki, A. Philalithis, Spatial-Mathematic methods for analysis of indicators of mortality, IJAEST, 2 (2010) 135:14.

[11] H.J. Miller and E.A. Wentz, Representation and spatial analysis in geographic information systems, Annals of the Association of American Geographers, 93 (2003) 574-594.

[12] J.L.M. Christopher and A.D. Lopez, Mortality by cause for eight regions of the world: Global Burden of Disease Study, The Lancet, 349 (1997) 1269 – 1276.

[13] K. Kamran, S.J.N. McNabb, Z.A. Memish, R. Eckhardt, Wei Hu, D. Kossowsky, J.Sears, J. Arino, A. Johansson, M. Barbeschi, B. McCloskey, B. Henry, M. Cetron, J. S. Brownstein, Infectious disease surveillance and modelling across geographic frontiers and scientific specialties, The Lancet, 11 (2012) 70313-9.

[14] L. Anselin L, The Moran scatter plot as an ESDA tool to access local instability and spatial associations, paper presented at the GISDATA Specialist meeting on GIS and Spatial analysis, Amsterdam, The Netherlands, December 1-5 (1993) (West Virginia University, Regional research Institute, research paper 9330).

[15] L. Anselin, Local indicators of spatial association – LISA, Geographical analysis, 27 (1995) 93-115.

[16] P.A.P. Moran, Notes on Continuous Stochastic Phenomena, Biometrika 37 (1950) 17–23.

[17] P. Gould and R. Wallace, Spatial structures and scientific paradoxes in the AIDS pandemic, Geografiska Annaler, 76 (2004) 105-116.

[18] R. Peto, J. Boreham, A.D. Lopez and M.T.C. Heath, Mortality from tobacco in developed countries: indirect estimation from national vital statistics, The Lancet, 339 (1992) 1268-1278.

[19] S.B. Bell, R.E. Hoskins, L.W. Pickle and D. Wartenberg, Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public, International Journal of Health Geographics, 5 (2006) 49.

[20] WHO database, World Health Organization, Available at: http://www.who.int/library/databases/en/ , Accessed in: 28/01/2012.

[21] WHO, World Health Organization, Fact sheet N°297, October 2011, Available at: http://www.who.int/mediacentre/factsheets/fs297/en/ , Accessed in: 28/01/2012.