SPATIAL STATISTICAL METHODS IN THE ANALYSIS OF PUBLIC HEALTH DATA

Prof. Assoc. Dr. Bederiana Shyti
MSc. Elona Fetahu
Department of Mathematics, Faculty of Natural Sciences, UE
PhD Elvira Fetahu
Department of Marketing and Engineering, Economic Faculty, UE

Abstract

Various studies claim that cancer is likely to be caused by the diverse environmental pollutants; lifestyle, i.e., poor diet, smoking, alcohol, stress, sun exposure, lack of physical activity, non-healthy weight; genetic inheritance; some kind of infections, etc. In most cases, around 90-95% of cancers are due to lifestyle and ecological factors that influence the living organisms. This implies that people affected by cancer most probably are clustered around the most polluted regions, meaning that the geographical location has an effect in the chances of contracting the disease. We use spatial statistical methods to support this and to point out the contrasts in rates that come out from different geographical distributions of the population. After the heart disease, cancer is the second cause of the worldwide deaths, in spite of the intensive research done in the last years. Based on the information on the causes of deaths by group diseases, provided by the Albanian Institute of Statistics INSTAT for the last two decades, this fact is also true in the case of Albania, where on the first place we have the circulatory system diseases with an average of 242 deaths per 100.000 inhabitants a year, followed by an average of 78 deaths per 100.000 inhabitants a year caused by neoplasm. In this study we take into consideration some specific geographical areas and see how the critical points do influence in the higher chance of being affected by cancer. We use correlation to show the relation between number of sick people and air pollution rates.

Keywords: Spatial statistics, correlation, public health, spatial epidemiology, environmental risk factors

Introduction

Spatial epidemiology is an intersection of a variety of fields requiring knowledge from epidemiology, statistics, public health, geography as the study of distribution, etc. Grouping and connecting entities with same characteristics is in our human nature. In science there exist statistical methods which study the geographical location of the entities, their relative distance to each other, or their clusters. Coming to public health, these spatial statistical methods give important information on how a disease is spread, which are the territories affected by the disease and predict the next territories which have a higher possibility to be touched in order to control it. Studying the geospatial public health data, despite giving us information on where the disease is spread, it also provides us data on why the disease is found there. Several researches have concluded that a specific illness is found on some particular geographical location due to the characteristics of this location, which are the cause of the disease or contribute in its development. (Snow, 1855) Waldo Tobler in his first law of geography "Everything is related to everything else, but near things are more related than distant things", claims that adjacent entities are more alike to each other than the ones taken at further distances. (Tobler, 1970)

However, being exposed to the same features of the location, does not necessarily mean that someone will be affected by the disease. This indicates that, although being under alike conditions, for each individual, there exist a different probability to contract the disease within a specified period of time. The analysis of public health data uses spatial statistical methods to identify clusters or differences coming from different geographical locations. Our aim is to find a relation between groups of people being introduced to different levels of some component which has an influence on the person's risk to get the disease. This component might be the air pollution rate, water pollution, land pollution, etc. Some other factors influencing a person's risk to get the illness are age, gender, lifestyle factors and socioeconomic status. When we notice that some certain event is happening to some certain population within a specified timeframe, we get indications to start studying the disease. Researcher's job is to estimate the risk factor, which is always changing. (Waller & Gotway, 2004)

Methodology

In our study we will focus on cancer, the second cause of human deaths worldwide, after the cardiovascular diseases. Cancer is caused on its 20% of cases by tobacco use, 20 % by viral infections like hepatitis B, hepatitis C and papillomavirus, 10% because of obesity, unhealthy diet, lack

of physical activity, 5-10% by inherited genetic mutations. Some other factors are ionizing radiation[1] and environmental pollutants.

Firstly, we do some review and collect the latest data available related to cancer worldwide and specifically to Albania. Among men the 5 most common types of cancer for 2012 were lung, prostate, colorectum, stomach and liver cancer. Among women the 5 most common types of cancer were breast, colorectum, lung, cervix and stomach cancer. (World Health Organization, 2014) According to WHO's report on estimated cancer incidence and prevalence, in Albania the 10 most common types of cancer are as shown in Figure 1.

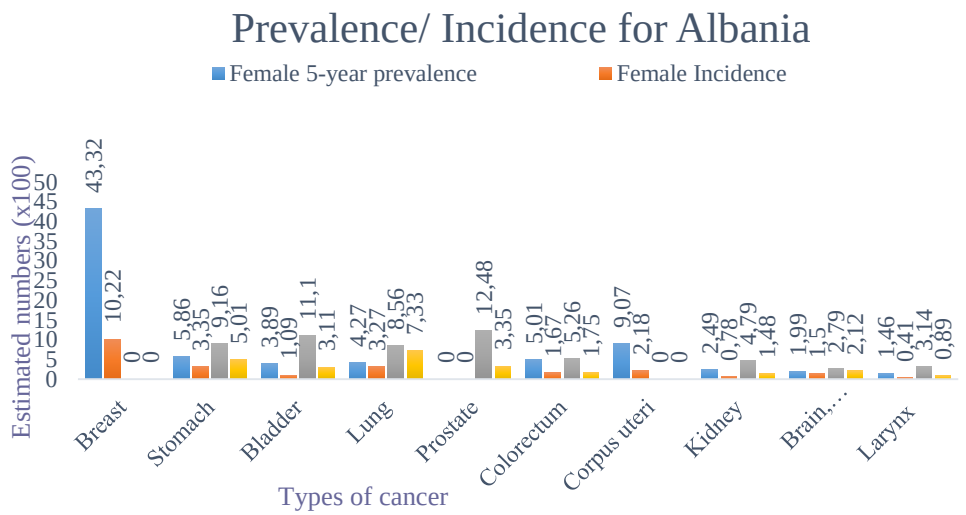


Figure 1 Prevalence/Incidence for Albania

Among various types of cancer, we turn our attention to lung carcinoma, the one that causes the biggest amount of cancer-related deaths worldwide, 1.589.925[2] deaths a year with the highest crude rate[3] of 22,5 per 100.000 a year. The vast majority of cases, i.e., 85-90%, are related to cigarette smoking. (Lung Carcinoma: Tumors of the Lungs, 15 August 2007) (Longo, et al., 2012) Smokers are 15 times more likely to die than nonsmokers, but lung cancer in the 15-20% of cases is also likely to happen to people who are second-hand smokers or who have never smoked. (Thun, et al., 2008) Genetic inheritance, exposure to carcinogens, and some forms of air pollution, although by a small amount and far less than smoking, have

[1] Harmful to living beings, but contributes in radiation therapy for the treatment of cancer.
[2] Latest WHO estimation on world cancer mortality for 2012, all ages: both sexes. (2014)
[3] Total number of events occurring among the popullation of a given geographical area over a specified period of time.

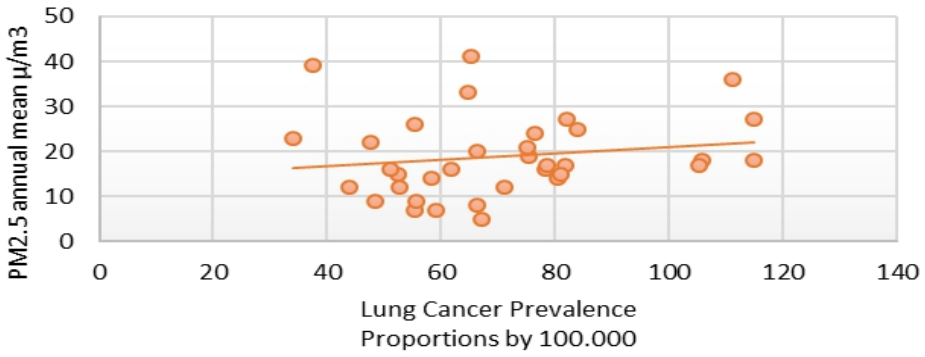
their own impact in increasing the chances of getting sick. (Mason, Broaddus, Martin, & King, 2010) (Lung cancer risk and causes, 2014)

Second, we look at some researches done on how air pollution rate affects the chance of being contracted by lung cancer. The particulate matters suspended in the air, which are due to natural causes or man-made, have a huge effect in our lives. As an evidence to this we can mention a study done in 2007, where it was found out that being exposed to $PM_{2.5}$ (fine particular matters with a diameter of 2.5 micrometers or less), increases the mortality caused by lung cancer by 15-21% per a 10μg/m^3 increase. (Chen, Goldberg, & Villeneuve, 2011) Next evidence comes from a later study done in 2013, involving 312.944 people from 9 European countries. It was observed that an increase by 10μg/m^3 in PM_{10}, rose the lung cancer rate by 22%. Even more, an increase by 10μg/m^3 in $PM_{2.5}$, rose the lung cancer by 36%. (Raaschou-Nielsen, Andersen, Beelen, & Samoli, 2013) Particulate matters found in the air are dangerous, because of their small size and the ability to get deep into the lungs, even more into the blood stream, causing lung cancer, cardiovascular illnesses, respiratory infections, DNA mutations and other serious health problems.

Based on the above mentioned facts, we turn to the World Health Organization's database on global air pollution[4]. They estimated the rates of $PM_{2.5}$ and PM_{10} from 1600 cities in 90 countries of the world. According to WHO the country with the highest outdoor pollution for 2014 is Pakistan with an average of $PM_{2.5}$ of 101μg/m^3, followed by Qatar 92μg/m^3, Afghanistan 84μg/m^3, Bangladesh 79μg/m^3, Iran 76μg/m^3, Egypt 74μg/m^3, Mongolia 64μg/m^3, United Arab Emirates 61μg/m^3, India 59μg/m^3 and Bahrain 57μg/m^3. There exist natural reasons and man-made reasons for making these countries the most polluted in the world. The reasons vary from country to country according to their development rate: high rate of construction, busy air traffic, growing number of cars, factories, burning coal and wood for cooking and heating, burning tires and plastic bags for fuel, dust, geographical limitations, industrial emission etc.

To sum up, we make a review on the cancer situation, world air pollution rates and studies done concerning the relation between these two components. By possessing the incidence/ prevalence/ mortality rate caused by lung cancer and also the pollution rates, using statistics, we will explain how these two elements relate.

[4] Ambient Pollution Database, WHO, May 2014



We use correlation (Mukaka, 2012) to show a linear connection between these two variables: annual mean of $PM_{2.5}$ [5]and pulmonary carcinoma prevalence[6] in 35 European countries. We find out that the correlation coefficient is 0.2, which is considered to be a low positive correlation, however shows that the $PM_{2.5}$ has an influence on the number of prevalence. Of course we should not expect to have a high correlation, due to the fact that air pollution is not the only one that determines the risk to get lung cancer, there is a set of reasons that influence on this.

In our case study we will concentrate on Albania, specifically the city of Elbasan.

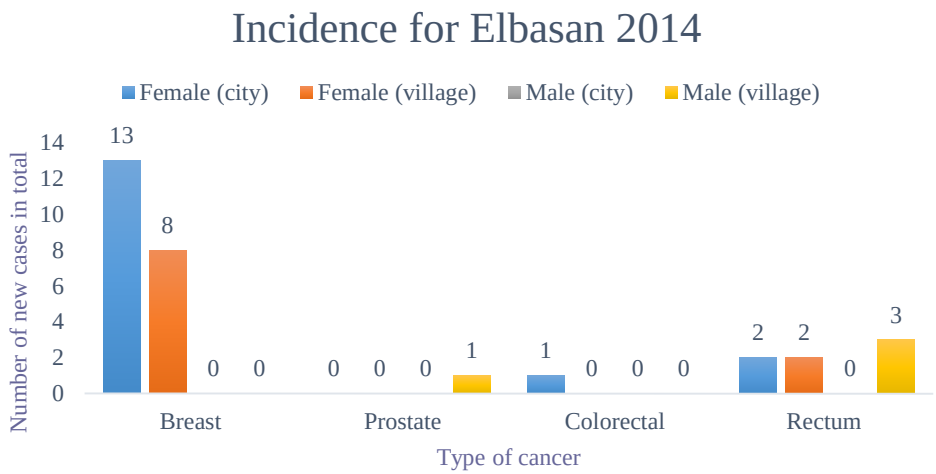


Figure 2 Incidence for cancer cases Elbasan, age 25-65 (2014)

[5] World Health Organization Air Pollution Database May, 2014.

[6] World Health Organization Estimated Cancer Incidence, Mortality and Prevalence Worldwide Database, 2012.

Case study

As indicated, here we will focus on Elbasan's lung cancer cases and its most polluted geographical areas. Below we will mention the reasons why it is important to research on this city and the reasons why the scientific research work faces limitations. The city of Elbasan, with a population of 78.703 inhabitants located in the Elbasan County with a population of 295.827 inhabitants[7], once considered to be the market garden city of Albania, today is one of the most polluted. (Albanian Institute of Statistics, 2010) (Brown, 2004) The main causes for this are due to the modes of transportation, release of dust and gases which are being spewed out by the construction companies and industrial businesses. 30 chimneys giving out smoke and dust, burned waste, increased number of construction, damaged roads and traffic have influenced a lot in the air and land quality. To learn about Elbasan's air pollution rate, we refer to the database of WHO. Unfortunately, on the database we cannot find information related to the air pollution of Albania, even more related to the city of Elbasan, but we find several reports when it comes to this topic. Some of them claim that the air pollution in the industrial areas of Elbasan is within the Albanian norms of pollution, although it is higher in comparison to the EU norms. (Elbasan, pollution within domestic norms, 2013) (Municipality of Elbasan, 2011) Some other claim that the pollution level is within the EU norms. (Smog in Elbasan, 2012) According to the Municipality of Elbasan, in 2007, some air pollution measurement equipments were installed in three different points of the city: near the municipality, near the former directorate of the metallurgic plant and near the "Aleksander Xhuvani" University, out of which only the one near the municipality is currently working. According to these measurements it was observed that the amount of the PM_{10} is decreasing from year to year (from 66.24 $\mu g/m^3$ in 2008 to 56.16 $\mu g/m^3$ in 2009) and that it is lower than the norm defined for Albania (70 $\mu g/m^3$). Anyhow on some specific days these values turn out to be up to two times higher than the norm, namely 136.34 $\mu g/m^3$ for 2008 and 153.35 $\mu g/m^3$ for 2009. Taking into consideration that the measurement equipments near the former directorate of the metallurgic plant are not working, these conclusions are not really reliable. (Municipality of Elbasan, 2011) According to the Official Journal of the Republic of Albania's[8] report for 2014, neither the Institute of Public Health (ISHP[9]), nor the National Environmental Agency (AKM[10]) have implemented the necessary equipment for the assurance of the data or

[7] Population and Housing Census 2011 - page 84
[8] Fletorja Zyrtare e Republikes se Shqiperise
[9] Instituti i Shendetit Publik
[10] Agjensia Kombetare e Mjedisit

have carried out the appropriate calibration of the air pollution measuring instruments. The latest published data from the Environmental Air Quality (CAM[11]) claim that the annual average of PM_{10} in the city of Elbasan was 79.24 μg/m^3 and 88 μg/m^3 for the years 2011 and 2012 respectively, while the EU standard for PM_{10} is 40 μg/m^3. This clearly shows that the city of Elbasan is facing big problems with the air pollution. (Për miratimin e Strategjisë Kombëtare për Cilësinë e Ajrit të Mjedisit, 2014) Its hospital statistics show that every day 2 people die, from which only one from natural causes. (Rukaj, 2012) According to the Albanian Population and Housing Census for 2011, 78.703 inhabitants live within the city of Elbasan and 109.959 inhabitants live in the municipalities of the district of Elbasan. (Instituti i Statistikes Republika e Shqiperise, 2011) The Statistics Office from the Department of Public Health (DSHP[12]) of the district of Elbasan provided us with the number of people affected by cancer coming from the city and villages around. For 2014, the incidences are illustrated in Figure 2.

Limitations

We had to use the division: center of the city and villages around, as there is no database showing the number of incidences coming from each neighborhood, nor data showing the air pollution rates of the neighborhoods, or at least the pollution rates of the most contaminated geographical areas. On the other side, in order to complete the picture of the situation in Elbasan, it is necessary to have a wider range of years. Yet, even when the information exists, it is hard to get access to it, because of the fact that datum are not being processed, neither published, making them not accessible to all.

As a further direction, if we work on in order to form a geocoded database of the most polluted points of the Elbasan district and number of sick people around these zones, we can use bivariate K-function (Dixon, El-Shaarawi, & Piegorsch, 2002) statistical method to estimate the spatial dependence of sick people around the critical regions. (Bryn, et al., 2005)

It would be of big interest to study on the cities of Albania too, so we could have a clearer picture of the situation. For this we need to know each city's air pollution rate and number of incidence/ prevalence/ mortality.

Conclusions

From our analysis on 35 European countries we reinforce the claim on the existence of a positive relation between air pollution rate and number of people who suffer from cancer. On our next step we aimed to implement the same research on the case of the district of Elbasan, which is considered

[11] Cilesia e Ajrit te Mjedisit
[12] Departamenti i Shendetit Publik

to be one of the most polluted in Albania. Although many people die from cancer in this district, and the air pollution rates are considered to be high, we noticed that there is not much of official information available. Given that the Department of Public Health cannot issue data on the number of incidences coming from each neighborhood, as a further suggestion, it may be considered to turn to the Regional Health Centers of the city. Regarding the pollution, it's crucial to pay attention to the equipments needed in order to measure the rates and make the data public. It is important to stay updated with the public health data, to work on and process them, in order to grow our awareness on how to look after the place we live in, the way we live in, and to take safety measures on how to improve the situation in order to prevent further incidences of illnesses.

References:

Albanian Institute of Statistics. (2010). *Popullsia sipas rretheve 2001-2010.*

Anand, P., Kunnumakara, A. B., Sundaram, C., Harikumar, K. B., Tharakan, S. T., Lai, O. S., . . . Aggarwal, B. B. (2008). Cancer is a Preventable Disease that Requires Major Lifestyle Changes. *Springer Open Choice*, 2097-2116.

Brown, P. (2004). Elbasan: from market garden to dead zone. Retrieved from http://www.theguardian.com/environment/2004/mar/27/internationalnews.pollution

Bryn, A. S., Melly, S. J., Sanchez, B. N., Aarti, P., Stephen , B., & Steven, G. L. (2005). Clustering of Fast-Food Restaurants Around Schools: A Novel Application of Spatial Statistics to the Study of Food Environments.

Chen, H., Goldberg, M. S., & Villeneuve, P. J. (2011, May 11). A Systematic Review of the Relation Between Long-term Exposure to Ambient Air Pollution and Chronic Diseases. *23*(4), 243-298. doi:10.1515/REVEH.2008.23.4.243

Dixon, P. M., El-Shaarawi, A. H., & Piegorsch, W. W. (2002). Ripley's K function.

Elbasan, pollution within domestic norms. (2013, 04 12). Retrieved from http://top-channel.tv/english/artikull.php?id=8691

Instituti i Statistikes Republika e Shqiperise. (2011). *Population and Housing Census.* Retrieved from http://www.instat.gov.al/media/195817/4__elbasan.pdf

Longo, L. D., Fauci, A. S., Kasper, D. L., Hauser, S. L., Jameson, J. L., & Loscalzo, J. (Eds.). (2012). *Harrison's Principles of Internal Medicine* (18 ed.). McGraw-Hill.

(2014). *Lung cancer risk and causes.* United Kingdom: Cancer Research UK. Retrieved from http://www.cancerresearchuk.org/about-cancer/type/lung-cancer/about/lung-cancer-risks-and-causes

Lung Carcinoma: Tumors of the Lungs. (15 August 2007). *Merck Manual Professional Edition, Online edition*.

Mason, R. J., Broaddus, V., Martin, T., & King, T. E. (2010). *Murray and Nadel's Textbook of Respiratory Medicine* (5th ed.).

Mukaka, M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *24*.

Municipality of Elbasan. (2011). Elbasan deserves even more: Strategic development plan of the city 2010-2020. Elbasan.

Për miratimin e Strategjisë Kombëtare për Cilësinë e Ajrit të Mjedisit. (2014, September 22). Tirana, Albania: Fletorja Zyrtare e Republikes se Shqiperise.

Raaschou-Nielsen, O., Andersen, Z. J., Beelen, R., & Samoli, E. (2013, July 10). Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *14*, 813-822. doi:10.1016/S1470-2045(13)70279-1

Rukaj, M. (2012). Elbasan, the polluted city. Retrieved from http://www.balcanicaucaso.org/eng/Regions-and-countries/Albania/Elbasan-the-polluted-city-124037

Ruktanonchai, C. W., Pindolia, D. K., Striley, C. W., Odedina, F. T., & Cottler, L. B. (2014). Utilizing spatial statistics to identify cancer hot spots: a surveillance strategy to inform community-engaged outreach efforts. *International Journal of Health Geographics*.

Smog in Elbasan. (2012). Tirana. Retrieved from http://www.top-channel.tv/english/artikull.php?id=7443

Snow, J. (1855). *On the Mode of Communication of Cholera.*

Thun, M. J., Hannan, L. M., Adams-Campbell, L. L., Boffetta, P., Buring, J. E., Feskanich, D., . . . Samet, J. M. (2008). Lung Cancer Occurrence in Never-Smokers: An Analysis of 13 Cohorts and 22 Cancer Registry Studies.

Tobler, W. R. (1970, June). A Computer Movie Simulating Urban Growth in the Detroit Region. *46*, 234-240. Retrieved from http://www.jstor.org/stable/143141

Waller, L. A., & Gotway, C. A. (2004). *Applied spatial statistics for public health data.* Hoboken, New Jersey: John Wiley & Sons, Inc.

World Health Organization. (2014). *Cancer.* Retrieved from http://www.who.int/mediacentre/factsheets/fs297/en/