SPATIAL ANALYSIS OF THE RESULTS OF HEALTH SCREENING IN AFYONKARAHISAR

Saffet Erdoğan¹, Dilek Toprak², Levent Ö zgür³,

Abstract

This study aimed to explore GIS aided spatial analysis of distribution of results of, hypertension and diabetes screening between the towns of Afyonkarahisar provinces of Turkey and identify cluster areas to implement precautionary measures and provisions by health agencies for public health. In this study health data have been taken from the health screening. Empirical Bayes smoothing is used to remove background noise from the raw disease rates because of the sparsely populated cities and small number cases of cities. To detect global variations and trends in the values of smoothed rates over the neighboring towns, spatial rate smoothing based on spatial moving areas technique is performed. Cluster analyses were performed whether the cases show clustering or located closer by chance. Geostatistical analyses performed to determine the problematic areas.

INTRODUCTION

Diabetes Mellitus type 2 is a serious and common health problem all over the world and is mostly associated with hypertension, obesity, metabolic syndrome and cardiovascular disease (CVD). It is more prevalent in some parts of the world and in some ethnic groups. Several national surveys have suggested a higher prevalence of Diabetes Mellitus (DM) in developing countries.¹ Both forms of diabetes may be inherited in genes, but it is known that in type 2 DM genes are more important. A family history of diabetes can significantly increase the risk of developing diabetes. The importance of the disease is not only for its worldwide high prevalence but also for its complications, high cost of treatment, mortality and morbidity tendency. Prevalence of unknown diabetes increasing considerably which means many people will be affected in the future without being aware of their still prediabetic position. Another important chronic disease of the world that screened in this study is hypertension. HT is one of the most important modifiable risk factor for blindness, CVD and renal disease in Western and Asian populations.²,³ The prevalence of hypertension (HT) in Egypt was estimated to be 26.3%, 33.7% in Korea, 27.2% in China and 24% in United States.²,⁴,⁵,⁶,⁷ Several national surveys have suggested a higher prevalence of HT in developing countries that is nearly 33.7% for Turkey.²,⁵,⁸ Although the very important advances in treatment, control of HT is still not satisfactory in our region. Many studies identified that only 44.7% of hypertensive patients were aware of their diagnosis, only 28.2% were taking prescribed medication, and only 8.1-28.8%, who were taking antihypertensive medication, had a blood pressure (BP) <140/90 mm Hg.²,⁹ These data indicates that many patients still under high risk.

¹ Assist. Prof. Dr. Afyon Kocatepe University, Faculty of Engineering, Surveying Engineering Division. saffet_erdogan@hotmail.com, serdogan@aku.edu.tr
² Assist. Prof. Dr. Afyon Kocatepe University, Faculty of Medicine, Family Practice Division
³ Research Assistant. Afyon Kocatepe University, Graduate School of Natural and Applied Science
As a result, DM, and HT diseases are important and prevalent problems also for our country. There are different factors such as, lifestyle, nutritional habitus, environment and partially genetic factors affect these diseases related to geographical distribution of resident. Meanwhile, medical geography is relatively a new concept in Turkey, and the sheer size of our country, varied life styles, climatic zones and environmental condition. Therefore, we aimed to investigate: prevalence of HT (previously diagnosed or the first diagnose) by blood pressure measurement and history of HT; and prevalence of type 2 DM among counties of Afyonkarahisar province as taking a part of a big epidemiological research made by Sozbilir et al. by using GIS and spatial analysis.

**DISEASE MAPPING AND SPATIAL ANALYSES**

Disease maps have been playing a key descriptive role in epidemiology, providing rapid visual summaries of geographic information and may identify patterns in the data that are missed in tabular presentations. These maps are useful tools for many purposes such as: to identify regions of unusually high risk in order to take preventive measurements; to provide reliable maps of disease risk in a region to allow better resource allocation and risk assessment. Recently, geographical information system (GIS) technology has been used as an innovative and important tool for visualization of disease data in epidemiology. GIS may also involve sophisticated spatial analyses of diseases that contributing environmental factor. Spatial analysis refers to the ability of the analysis to manipulate spatial data into different forms and to extract additional meaning as a result. It involves a variety of methods and procedures developed in different disciplines. Typically, the methods used in these analyses can be grouped in three parts; methods for point pattern analysis, methods for areal data and geostatistics. Of these, the geostatistical approach is most relevant to epidemiological analysis conducted at the landscape scale. Geostatistics focuses on the analysis of spatially distributed variables and the prediction or estimation of values at unsampled locations.

**STUDY AREA AND DATA COLLECTION**

The study was conducted in Afyonkarahisar, a less development middle Anatolian province with the population of 812,416, between November 2005 and February 2006. Large parts of this province are predominantly rural. The present study was approved by the Afyon Kocatepe University Faculty of Medicine Clinical Research Ethics Committee and written informed consent was obtained from all participants. In this community based cross sectional study, a total of 2035 people, from 75 different screening regions (18 urban, 57 provinces) were detected according to the population records of the year 2000, which represent the population of the area appropriately. Locations of the screened villages are shown in Figure 1. A total of 7000 km. roadway was driven for the research by a team of 15 physicians, 1 nurse and a driver. The records of the regional health institutions were used in order to determine the subjects. People older than 18 years old were grouped as 19-40 years old, 41-64 years old, 65 and over aged. According to population distribution of year 2000, we determined the minimum number of people of each group with the 2% margin of error (Table 1). Distribution of screened people according to gender and age group are shown in Table 1. Planning total number of 1998 was become as 2035 in the screening. The study group selected randomly from the “Family Cards” of the primary health centers, regarding the gender and ages. Only one person selected from every house. The subjects were informed about the study by telephone
interviews one night before, their approvals were obtained, and their transport to the health institutions, where the study would be conducted, was provided. The data were collected by a questionnaire in which face to face survey method was performed by the physicians. The subjects were divided into three groups according to age (19-40, 41-64, 65 and over aged groups). These study groups were evaluated by sociodemographic features medical history, physical examination, blood (including glucose levels) and urine analysis.

As this study is a part of a big epidemiological research, only the data about type 2 DM and HT were regarded. Type 2 DM was recognized according to World Health Organization (WHO) criteria. People without previously diagnosed diabetes were categorized according to WHO diagnostic criteria as fasting blood glucose was > or =126 mg/dl. Total number of patients who have type 2 DM was calculated by regarding fasting glucose (126mg/dl and over), blood glucose level ≥200mg/dl at any time of day and history of DM. In the study, population patients with known diabetes, including subjects receiving oral antidiabetic agents or insulin at the time of the study or some with high fasting glycemia not receiving any treatment except a diet were accepted as DM. In addition, the subjects who prior to the study had not been receiving hypoglycemic agents or in whom fasting glycemia had been over 126 mg/dl were calculated as diabetic patients. The subjects whose body mass index (BMI)≥25 kg/m^2 were accepted as obese.

Table 1. Distribution of planned and screened people according to gender and age group

<table>
<thead>
<tr>
<th>Urban</th>
<th>Population Planning Number of Screened persons according to 2% margin of error</th>
<th>Number of Screened Female</th>
<th>Number of Screened Male</th>
<th>Number of Screened Total Persons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afyonkarahisar</td>
<td>201110</td>
<td>83</td>
<td>137</td>
<td>19</td>
</tr>
<tr>
<td>Başmakçı</td>
<td>15084</td>
<td>2</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>Bayat</td>
<td>8753</td>
<td>5</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Bolvadin</td>
<td>7988</td>
<td>31</td>
<td>39</td>
<td>12</td>
</tr>
<tr>
<td>Çay</td>
<td>45635</td>
<td>15</td>
<td>49</td>
<td>6</td>
</tr>
<tr>
<td>Çobanlar</td>
<td>12364</td>
<td>15</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Dazkiri</td>
<td>15620</td>
<td>13</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>Dinar</td>
<td>88304</td>
<td>38</td>
<td>83</td>
<td>19</td>
</tr>
<tr>
<td>Emirdağ</td>
<td>47396</td>
<td>17</td>
<td>34</td>
<td>7</td>
</tr>
<tr>
<td>Evciler</td>
<td>9486</td>
<td>3</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Hocalar</td>
<td>12824</td>
<td>1</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>Ihsaniye</td>
<td>33220</td>
<td>10</td>
<td>47</td>
<td>3</td>
</tr>
<tr>
<td>İcibhisar</td>
<td>21978</td>
<td>14</td>
<td>19</td>
<td>2</td>
</tr>
<tr>
<td>Kizilören</td>
<td>4132</td>
<td>3</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Sandikli</td>
<td>76618</td>
<td>33</td>
<td>84</td>
<td>14</td>
</tr>
<tr>
<td>Sincanli</td>
<td>58536</td>
<td>13</td>
<td>80</td>
<td>14</td>
</tr>
<tr>
<td>Sultandağı</td>
<td>22184</td>
<td>9</td>
<td>26</td>
<td>3</td>
</tr>
<tr>
<td>Şuhut</td>
<td>59284</td>
<td>26</td>
<td>47</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>812416</td>
<td>331</td>
<td>728</td>
<td>136</td>
</tr>
</tbody>
</table>

**METHODS**

Rate Smoothing
Annual mid-year population estimates for the period were used as denominators for the calculation of crude rates. Since many villages in Afyonkarahisar are rural and sparsely populated, a small number of cases in such villages may produce high incidence rates that do not reflect a true disease cluster. In order to overcome the problem of rate instability, various smoothing methods are usually employed. The idea in smoothing is to borrow the information from other small areas for the estimation of the relative risk. We used empirical Bayes smoothing in this study. Empirical Bayes smoother uses Bayesian principles to guide the adjustment of the raw rate estimate by taking into account information in the rest of the sample. The principle is referred to as shrinkage, in the sense that the raw rate is moved towards an overall mean, as an inverse function of the variance.

Figure 1. Locations of screened villages in the province

Geostatistical Analyses
Analyses of the geographical distribution of the incidence of diseases and their relationships to potential risk factors have an important role in public health studies. Traditionally, choropleth maps or aggregated area maps are commonly used to display the incidence rates. These maps are simple to construct, requiring little statistical expertise. However, choropleth maps must be interpreted with caution, as the area borders are arbitrary, varying shape and size of areas may causes misinterpretations, and they imply a constant rate over the region at the borderline. Continuous maps, presenting smoothed estimates of disease rates, may strengthen the ability of communicate event patterns visually by avoiding artificial administrative
In this study, Ordinary Kriging (OK) was used for estimation of spatial distribution of smoothed prevalence rates of diseases.

In order to evaluate the results obtained after kriging, how well the experimental semivariogram was fitted to the theoretical one, cross-validation technique was performed. The cross validation method involves using all of the raw data for comparison. The main advantage of the method is a clearly defined and user independent. To define the accuracy of an interpolation method statistically, the difference between the raw data and the value obtained from interpolation method needs to be calculated via validation methods. Five error statistics of predictions, the mean error (ME), the root mean square error (RMSE), the average standard error (ASE), the mean standardized error (MSE) and the root mean square standardized error (RMSSE), were used for the cross-validation analysis as follows.\(^{21}\)

\[
ME = \frac{1}{N} \sum_{i=1}^{N} \{z(x_i) - \hat{z}(x_i)\}
\]

\[
MSE = \frac{1}{N} \sum_{i=1}^{N} \frac{z(x_i) - \hat{z}(x_i)}{\sigma(i)}
\]

\[
ASE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \sigma(i)}
\]

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \{z(x_i) - \hat{z}(x_i)\}^2}
\]

\[
RMSSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\frac{z(x_i) - \hat{z}(x_i)}{\sigma(i)}\right)^2}
\]

where; \(\hat{z}(x_i)\) is the predicted value, \(z(x_i)\) observed value, \(N\) is the number of values, \(\sigma\) is standard error for location i.

**RESULTS**

Arc GIS 9.2 and GeoDa 0.9.5-I softwares were used for visualization and spatial analyses of the disease data in the study. Empirical Bayes (EB) smoothing is calculated via EB tool created by National Cancer Institute of USA in ArcGIS 9.2 and raw incidence rates were replaced with their globally smoothed values. Figure 2 shows choropleth maps of the crude rates and EB smoothed rates of total type 2 DM and HT prevalences according to cities of Afyonkarahisar. Each map is a choropleth map where the natural break method for classification of the data has been applied to reflect the distribution best. The smoothed rate maps showed no clear spatial pattern of trend or clustering. Therefore, spatial rate smoothing based on the notion of a spatial moving average was used to emphasize any global variations and trends in the data. Afterward, the spatial rate smoothed incidences were computed from the total number of cases in a spatial window divided by the total number of people at risk within that window, which was specified using a spatial weights matrice, which using nearest 6 neighbors for contiguity including both city and its neighbor cities locations.

To explore spatial dependence, showing how the disease rates were correlated in the country, Moran’s I values were calculated. Moran’s I is produced by standardizing the spatial auto covariance by the variance of the data using a measure of the connectivity of the data. The range of possible values of Moran’s I is taken -1 to 1
since positive values indicate spatial clustering of similar values and negative values indicate a clustering of dissimilar values. Meanwhile, to assess the risk of diseases in each city, excess risk maps were produced too. The excess risk is the ratio of the observed rate to the average rate computed for all diseases. This average is not the average of the city rates, but calculated as the ratio of the total sum of all cases over the total sum of all populations at risk. Type 2 DM and HT did not show strong evidence of clustering. Başmakçi city had the highest excess risk value of 1.85 with type 2 DM prevalence and 1.78 with HT prevalence.

Figure 2. Choropleth maps of diseases’ prevalences according to towns (left side: raw rates; right side: smoothed rates)

Choropleth maps are effective in showing the variation in regional data but are also known to have problems associated with them as the uneven shape and size of the different regions produce a visual bias. The geostatistical prediction method of OK technique was performed to generate continuous prevalence maps according to gender and age of groups for all diseases. In order to generate maps from smoothed data via OK, spatial dependence was modeled by semi-variograms. The variance of rates between pairs of villages separated by a certain distance is plotted against their respective distances in an empirical semivariogram. Different models were fitted to this scattergram with nugget effect. The specification of nugget effect makes allowance for measurement error at a village. This avoids the prediction honoring every observation, which would result in sharpness in a map. The best-fit variogram
model parameters were optimized via cross validation with 12 lags and a lag size of 16 km. Model parameters for all disease rates are shown in Table 2.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Transformation</th>
<th>Model</th>
<th>Range</th>
<th>Partial Sill</th>
<th>Nugget</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT</td>
<td>None</td>
<td>Gaussian</td>
<td>24.9</td>
<td>0.00030489</td>
<td>0.0000023305</td>
</tr>
<tr>
<td>DM (type 2)</td>
<td>Box Cox (Parameter: 0.9)</td>
<td>Gaussian</td>
<td>166.5</td>
<td>0.0075729</td>
<td>0.0061959</td>
</tr>
</tbody>
</table>

Error predictions provide important information about the deficiencies of the methods. Statistical strength of OK method in the analysis of the spatial structure of data is the procedure yields predictions of standard errors or error variances for expected values. In order to compare the performance of the models, summary statistics of prediction errors calculated by cross validation method are shown in Table 3.

<table>
<thead>
<tr>
<th>Disease</th>
<th>ME</th>
<th>MSE</th>
<th>ASE</th>
<th>RMSE</th>
<th>RMSSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT</td>
<td>-0.0008555</td>
<td>-0.01918</td>
<td>0.01679</td>
<td>0.01438</td>
<td>0.9899</td>
</tr>
<tr>
<td>DM (Type 2)</td>
<td>-0.0006135</td>
<td>-0.009758</td>
<td>0.07059</td>
<td>0.08313</td>
<td>1.158</td>
</tr>
</tbody>
</table>

The ME and MSE indicate the degree of bias in model prediction and should be close to zero. The RMSE and ASE reveal the precision of prediction and should be as small as possible. The RMSSE compares the error variance with kriging variance, and should be close to 1. Interpolated maps of disease rates with equal interval classification are shown in Figure 3. After interpolation process, minimum, maximum, average and standard deviation statistics of disease prevalences according to age group and gender were calculated (Table 4, Table 5). Interpolation of the regional estimates overcomes the areal bias problem and generated isopleth maps were powerful to indicate the continuous values of disease prevalences. Başmakçı and Sultandağı region had highest values of HT prevalence. Başmakçı region had the highest type 2 DM prevalence confirming with the smoothed choropleth maps. Meanwhile using the Band Collection Statistics tool, we determined a correlation value of 0.301 between the prevalences of type 2 DM and HT.
Table 4. Descriptive statistics of prevalence rates according to gender

<table>
<thead>
<tr>
<th>DISEASE</th>
<th>MALE</th>
<th>FEMALE</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Max</td>
<td>Avrg</td>
</tr>
<tr>
<td>HT</td>
<td>0.0300</td>
<td>0.3327</td>
<td>0.1346</td>
</tr>
<tr>
<td>DM (Type 2)</td>
<td>0.1129</td>
<td>0.1133</td>
<td>0.1131</td>
</tr>
</tbody>
</table>

Table 5. Descriptive statistics of Prevalence rates according to age group

<table>
<thead>
<tr>
<th>DISEASE</th>
<th>AGE GROUP OF 19_40</th>
<th>AGE GROUP OF 40_64</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Max</td>
</tr>
<tr>
<td>HT</td>
<td>0.0694</td>
<td>0.1760</td>
</tr>
<tr>
<td>DM (Type 2)</td>
<td>0.0334</td>
<td>0.0610</td>
</tr>
</tbody>
</table>

DISCUSSION

It is known that prevalence of DM type 2 and HT increases progressively with age in the world. More than 20% of the populations aged over 60 have type 2 diabetes. This was also similar for our study as shown in Table 5, that in older age group the prevalence of the diseases was higher than the younger age group.

When obesity limit taken as BMI ≥ 25 kg/m² it was found that among nonobese subjects only 14.9% were hypertensive and 8.9% of them were diabetic. On the other hand, among the patients whose BMI were ≥ 25 kg/m² the prevalence of HT and DM type 2 were 27.96% and 14.89% respectively in the study. As most of the obese subjects were women, this may explain why the DM type 2 and HT were prevalent among women. In another point of view, among diabetic patients obesity was more prevalent in women (45.6%) than men (31.5%). This may explain the different distribution of DM type 2 in this region that is also strickly related with wrong diet and sedanter life style. Also in obese women, HT was more prevalent (41.9%) than obese men (24.0%).

While there is good evidence for a strong genetic contribution to both obesity and diabetes, the increase in these conditions in both developed and developing countries appears to be due to a changing balance between energy intake and energy expenditure through physical activity. Epidemiological data suggest that lifestyle changes involving increased physical activity and reduced energy intake will at least partially prevent type 2 diabetes. It is only very recently that prospective intervention studies have clearly confirmed the efficacy of such measures.

On the other hand, the annual incidence of DM type 2 in Turkey rises very rapidly, currently stands at 300.000 and hence, its prevalence rises correspondingly. This makes our study more important as we diagnosed unknown hypertension during the research. Regarding a study about DM epidemiology among 20 and over aged people in Turkey (TURDEP), DM prevalence is 7.2% (8% in women, 6.2% in men) Ozdemir et al found the prevalence as 6.4% in 30 and over aged group in another city (Sivas) of Turkey.

In a study by Yumuk et al. the crude impaired fasting glucose (IFG) rate was found as 24% (27.1% in women and 18.5% in men) and the diabetes rate 8.4% (8% in women and 9.1% in men) in Konya which is a very close city to our region. In the same study the overall prevalence of diabetes in those > or =20 years of age was 6.0% (n=160).

Başmakçı city, which had the highest excess risk rate with DM, is located in the south of province. The distribution of high prevalence of Afyonkarahisar may be due to diet problem of the people, which is mostly consisting of carbohydrate and fat. As we know that other than genetics many factors affect the precipitation of diabetes, like
diet, exercise, stress, obesity etc; we can explain the high distribution of DM in this region regarding these conditions.

In some studies the prevalence of hypertension was significantly higher in males (29.6%, 95% CI 28.3-31.0) compared with females (26.0%, 95% CI 25.0-27.1). On the other hand, some researches show a reverse or similar correlation of HT and gender. In a study, in Turkey, a higher prevalence in women than in men in 18 and over age group (36.1 versus 27.5%, P<0.001) was detected. Similarly, most of the hypertensive patients were women in our study, which is correlated with similar risk factors with DM type 2. Başmakçı city also had the highest excess risk rate with HT. Meanwhile, isopleths maps indicated a more specific region problematic with HT.

REFERENCES


