Association between esophageal cancer and drought in China by using Geographic Information System

Kusheng Wu⁎, Ke Li

Department of Preventive Medicine, Shantou University Medical College, Shantou 515031, PR China

Received 9 October 2006; accepted 3 January 2007
Available online 30 January 2007

Abstract

The objective of this ecological study was to discover associations between selected climate variables and esophageal cancer (EC) mortality in China using a Geographic Information System (GIS). A digital distribution map of EC mortality in China was established in GIS, which was based on one-tenth of nationwide population cause-of-death surveys conducted in mainland China in 1990–1992. Selected climate variables such as 30-year annual average precipitation and evaporation data of the sample areas were extracted from the environmental databases by zonal statistics finished in Spatial Analyst module of ArcInfo 9.0. Drought Indexes were calculated by using the precipitation and evaporation data and a digital distribution map of them was created to compare with the distribution of EC mortality. Correlation and regression analyses were applied to evaluate associations between the EC mortality rates defined at the sample areas and selected climate variables from the raster datasets. The results of the digital GIS maps of EC mortality and Drought Index show that the high EC mortality mostly occurred in areas with high Drought Index. Correlation and regression analyses also show weak negative correlation between precipitation and EC mortality (p < 0.001), and weak positive correlation between Drought Index and EC mortality (p < 0.001). This study presented a unique model for the link of cancer and climate using a GIS. The study suggests that drought plays a role in the occurrence and development of EC in China, however, other environmental, biological and genetic factors should not be ignored. There is need for further studies using multiple factors and more accurate and detailed environmental and health data.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Esophageal cancer; Climate; Drought; Ecological study; Spatial analysis; Geographic Information System

1. Introduction

Esophageal cancer is a common malignancy with high mortality, which is over 100 per 10^5 in some areas of China (Stoner and Gupta, 2001). It is generally recognized that EC is the result of multiple risk factors, such as environmental factors, biologic factors and genetic factors. Known risk factors for EC include lack of fresh fruits and vegetables, nutrient and micronutrient intake, infections of virus and fungi, nitrosamine, smoking and alcohol consumption, and so on (Stoner and Gupta, 2001). Although there are clear distribution belts of EC, associations between geographic factors, ambient climate and EC have been studied rarely, with only few studies conducted in Iran (Kmet and Mahboubi, 1972), China (Song, 1992), Kazakhstan (Akhtiamov and Kairakbaev, 1983) and Africa (Gendron et al., 1983). Akhtiamov and Kairakbaev (1983) reported that there was an inverse correlation between incidence of EC and altitude for relative region (p less than 0.001). High temperature and drought climate were reported possibly to influence the occurrence of fumonisin, which might be one of carcinogens of EC (Miller, 2001). But understanding the link of ambient climate and incidence and/or mortality of EC required more detailed and precise study.

In the past traditional papery maps were used to evaluate and describe the distribution of cancer, and provide many clues for cancer epidemiologic study. But information that traditional papery maps can provide is potentially limited. In the last decades, use of geographic mapping has been expanded especially since the development of GIS technologies which comprise data analysis through spatial strategies and a visual presentation of the information. The impact of GIS has been
widely felt in many fields that use geographic information — resource management, land-use planning, pollution and health, transportation and utilities, marketing, geosciences and planning of health care facilities.

The use of GIS has great potential for the management of chronic disease such as cancer and the analysis of clinical and administrative health care data. One of the more powerful features of GIS is the ability to link several databases such as geographic, clinical, and create a demonstration of the spatial distribution of the phenomenon. Therefore, the use of GIS in health related studies is emerging as an important and powerful analysis tool. GIS was used by researchers in implementing spatial analysis of cancer mortality in China (Lam, 1986), identifying populations potentially exposed to agricultural pesticides (Ward et al., 2000), estimation of long-term average exposure to outdoor air pollution (Hoeck et al., 2001), mapping and analyzing rates and distribution of child abuse in order to allocate special services (Emst, 2000) as well as many other applications.

This study reports the use of GIS in analyzing relationship between the mortality rate of EC and drought climate in the People’s Republic of China. The results of this study may provide an understanding of the effect of climate factors on EC, and suggest possible hypotheses about the mechanism linking exposure to drought condition and EC.

2. Materials and methods

2.1. Data collection and management

2.1.1. Digital maps

The 1:40000000 digital maps of China were provided by the State Bureau of Surveying and Mapping. The digital maps include layers such as national boundaries, provincial boundaries, county boundaries, residential areas, rivers, main railways and main highways. The one-tenth of population sample areas digital polygon maps were created in ArcInfo GIS software (ESRI, Inc., Redlands, CA) by selecting geocode from the existing national fundamental digital polygon maps were created in ArcInfo GIS software (ESRI, Inc., Redlands, CA) by selecting geocode from the existing national fundamental digital maps.

2.1.2. Esophageal cancer mortality data

The EC mortality data was obtained from one-tenth of nationwide population cause-of-death surveys conducted in mainland China in 1990–1992. The sampling method introduced in the surveys was cluster random sampling with two stages, stratified sampling and equiprobability (10%). Firstly, the study areas were stratified by province and city. Then all the counties/cities were ordered by their mortality rate from 1970s (There were also nationwide population cause-of-death surveys conducted in mainland China in 1973–1975, which included 2489 counties/cities of the whole nation.). Finally, 10% of the total counties/cities were selected in high, medium, and low mortality levels respectively. The EC mortality of the sample areas can thus be seen to represent the mortality of the whole nation. Because of difficulties, Xinjiang, Qinghai and Xizang province were not included in the surveys and 263 areas were sampled and surveyed finally. Since some areas were sections of a county (city), we selected 237 counties (cities) for our study by summing up. Standardized rates were computed by direct method using the Chinese population age distribution in 1982 as the standard of weights. The EC mortality database of the sample areas was created in Microsoft Excel according to male/female. Then a GIS for EC mortality was created in ArcInfo 9.0 based on the polygon map of the sample areas through inputting the mortality data regarded the county name as linked keyword. A dispersive digital polygon map of EC mortality value was drawn with graduated symbols. The EC mortality data was obtained from one-tenth of nationwide population cause-of-death surveys conducted in mainland China in 1990–1992. The sampling method introduced in the surveys was cluster random sampling with two stages, stratified sampling and equiprobability (10%). Firstly, the study areas were stratified by province and city. Then all the counties/cities were ordered by their mortality rate from 1970s (There were also nationwide population cause-of-death surveys conducted in mainland China in 1973–1975, which included 2489 counties/cities of the whole nation.). Finally, 10% of the total counties/cities were selected in high, medium, and low mortality levels respectively. The EC mortality of the sample areas can thus be seen to represent the mortality of the whole nation. Because of difficulties, Xinjiang, Qinghai and Xizang province were not included in the surveys and 263 areas were sampled and surveyed finally. Since some areas were sections of a county (city), we selected 237 counties (cities) for our study by summing up. Standardized rates were computed by direct method using the Chinese population age distribution in 1982 as the standard of weights. The EC mortality database of the sample areas was created in Microsoft Excel according to male/female. Then a GIS for EC mortality was created in ArcInfo 9.0 based on the polygon map of the sample areas through inputting the mortality data regarded the county name as linked keyword. A dispersive digital polygon map of EC mortality value was drawn with graduated symbols.

2.1.3. Climate databases

The environmental databases included 1961–1990 annual average precipitation data (ESRI raster format, 1:40000000) and 1961–1990 monthly and annual average evaporation data (Microsoft Excel format), obtained from Chinese Natural Resources Scientific Database. We used the 30-year average climate data to coincide with the retrospective EC mortality data, based on the reason that the occurrence and development of cancer is a long period and complicated process, and that our focus was on cancer and chronic exposure.

2.2. Spatial analysis

2.2.1. Extracting precipitation data of sample areas

The 30-year annual mean precipitation data of the 237 sample areas were extracted from the existing ESRI raster layer of precipitation through zonal statistics in GIS software. The sample areas digital polygon maps were used as zone dataset while the precipitation raster datasets layer was used as value raster, and all work was finished in Spatial Analyst module of ArcInfo 9.0.

2.2.2. Extracting evaporation data of sample areas

The evaporation database was created in Microsoft Excel, including 30-year monthly and annual average evaporation data, latitude and longitude coordinates of the 175 monitoring stations in mainland China. Then a digital point layer of the evaporation was created in ArcInfo 9.0 by using adding x, y data method (x, y represent latitude and longitude of the 175 monitoring stations). Inverse distance weighting (IDW) interpolation method was used to establish a continuous and smooth map of the evaporation data from the point layer so that we could extract the mean 30-year evaporation data of sample areas from it. We used IDW method due to lack of normality of distribution of evaporation and difficulty of transformation. IDW is a simple method for curve fitting, a process of assigning values to unknown points by using values from known points. A simple IDW weighting factor is

\[ w(d) = \frac{1}{d^2} \]

where \( w(d) \) is the weighting factor applied to a known value, \( d \) is the distance from the known value to the unknown value, and \( p \) is a user-selected power factor. Here weight decreases as distance increases from the interpolated points. Greater values of \( p \) assign greater influence to values closest to the interpolated point. In this study the value of \( p \) is 2.

When the map was established, we also extracted the 30-year annual mean evaporation data of the 237 sample areas from it through zonal statistics in GIS. The method and process were the same as extracting precipitation data. The sample areas digital polygon maps were used as zone dataset while the evaporation map layer was used as value raster, and all work was finished in Spatial Analyst module of ArcInfo 9.0.

2.2.3. Calculating Drought Index of sample areas

Drought Index is defined as ratio of annual mean potential evaporation to annual mean precipitation. We calculated Drought Index by using formula as follows:

\[ DI = \frac{e}{p} \]

where DI is the 30-year annual mean Drought Index, \( e \) is the 30-year annual average evaporation, \( p \) is the 30-year annual average precipitation. After they had been calculated, a digital distribution map of the Drought Indexes of the sample areas was drawn with graduated symbols by linking them with the digital maps made above in ArcInfo 9.0.

2.3. Statistical analysis

The fundamental hypothesis in this study is that the temporal and spatial variation in climate factors is associated with temporal and spatial variation in EC mortality. The relationship between EC mortality and climatic exposure was evaluated by correlation and regression analysis, and all statistical analysis was done with SPSS 13.0 (SPSS Inc., Chicago, Illinois, USA). The 30-year annual mean precipitation, evaporation and Drought Index were correlated with EC mortality. The Spearman’s correlation was used for analysis to observe how
strong relationship existed between climate factors and EC in mainland China. If there is significant correlation, then curve estimation was computed and we could select the best regression equation.

3. Results

We drew a dispersive digital polygon map of EC mortality value with graduated symbols in ArcInfo 9.0. Fig. 1 shows the distribution of EC mortality in mainland China except for Xinjiang, Xizhang and Qinghai provinces (no data).

The 30-year annual average precipitation data format is ESRI raster datasets at a scale of 1:4000000, which divides the land into approximately 1-square-mile rectangular units called sections. The raster atlas was gray-coded according to the value of precipitation (Fig. 2). Fig. 3 shows the spatial and smooth trend of the 30-year annual average evaporation predicted from the 175 monitoring stations using IDW interpolation method. When the precipitation and evaporation data of the sample areas were obtained, we could easily calculate the Drought Index of them in SPSS. The spatial distribution of the Drought Indexes of the sample areas is shown with graduated symbols (Fig. 4).
From the digital map of distribution of EC mortality, we can distinguish clearly the highest and lowest risk areas of EC in mainland China. When we compare the digital map of EC mortality with the digital map of Drought Index, we can see that the high EC mortality mostly occurred in the areas with high Drought Index.

Associations between precipitation, evaporation, Drought Index and EC mortality were evaluated using Spearman correlation analysis in SPSS 13.0. Table 1 shows weak negative correlation between precipitation and EC mortality, and weak positive correlation between Drought Index and EC mortality.

Regression analysis and curve fit were done in SPSS 13.0 after Spearman correlation analysis. The dependent variable (logarithm of EC mortality was used due to lack of normality of distribution of EC mortality data) contains non-positive values, and the minimum value is −0.21, so log transform cannot be applied. The Compound, Power, S, Growth, Exponential and Logistic models cannot be calculated for this
Drought Index 0.401

The dependent variable is logarithm of male EC mortality, the independent variable is Drought Index.

This study presents a unique model for the link of cancer and climate using a GIS. The maps that we created demonstrated the areas of high EC mortality, the areas which are short of precipitation, with strong evaporation and with high Drought Indexes. The information presented by the maps can be also presented in traditional ways such as tables produced from statistical analysis, but there is no doubt that these maps are a better way to present the phenomena.

From the result we can see that there is a relation between precipitation, Drought Index and esophageal cancer. But how shortness of precipitation and drought climate influence the occurrence and development of esophageal cancer? Before the prospect of anthropogenic climate change emerged, epidemiologists were not greatly interested in climate–health relations (McMichael et al., 2006). But drought appeared to interact with ecologic and socioeconomic conditions as well as life styles, magnifying the human impact of infectious diseases (Acuna-Soto et al., 2002). Modern epidemiology has focused mainly on studying risk factors for noncommunicable diseases in individuals, not ecological studies.

The characteristics of drought are expressed in terms of Drought Index, including intensity, duration and frequency. Drought is an elusive climate event and slow disaster because droughts develop slowly and have a prolonged existence, sometimes over many years. However the outcome of a drought related disaster could be wide diffused and devastating. It is one of the most adverse natural calamities which is called ‘life killer’. When there is drought, there is no drinking water for animals and for plant growth in the area. Some areas of China are water scarce regions under constant threat of drought, and the availability of water is an ongoing issue of struggle for the people. Basification and desertification are common features in most part of these areas. High salinity can promote the forming of nitrosamine, which is one of the important carcinogens of esophageal cancer (Lin et al., 2002). Plants and foods in these areas also contain more nitrosamine than other areas due to absorbing from water and soil. Nitrosamine can ultimately enter human body through food chain. The livelihood of the rural population is predominantly dependent on agriculture and livestock rearing, but in these drought areas agriculture is mainly rained and only the monsoon crop (kharif) is taken by farmers, so drought areas are relatively needy areas. In drought rural areas, the main crops are corn, wheat and kaoliang, so the rural people lack of fresh vegetables and fruits, which are protecting factors for EC. Drought is considered as a slow poison, no one knows when it comes, it can last many months in

### Table 1

<table>
<thead>
<tr>
<th>Climate factors</th>
<th>EC mortality (n=237)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td></td>
<td>Correlation coefficient</td>
<td>p</td>
</tr>
<tr>
<td>Precipitation</td>
<td>−0.223</td>
<td>0.001</td>
</tr>
<tr>
<td>Evaporation</td>
<td>0.072</td>
<td>0.278</td>
</tr>
<tr>
<td>Drought Index</td>
<td>0.401</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Equation*</th>
<th>Model summary</th>
<th>Parameter estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R²</td>
<td>F</td>
</tr>
<tr>
<td>Linear</td>
<td>0.058</td>
<td>14.57</td>
</tr>
<tr>
<td>Logarithmic</td>
<td>0.152</td>
<td>42.16</td>
</tr>
<tr>
<td>Inverse</td>
<td>0.180</td>
<td>51.65</td>
</tr>
<tr>
<td>Quadratic</td>
<td>0.124</td>
<td>16.58</td>
</tr>
<tr>
<td>Cubic</td>
<td>0.164</td>
<td>15.23</td>
</tr>
</tbody>
</table>

*The dependent variable is logarithm of male EC mortality, the independent variable is Drought Index.
a year or many years and its severity cannot be predicted. It is a part of the earth’s climate and occurs every year with no warning. Other climatic factors such as high temperature, high wind and low relative humidity are often associated with drought and can significantly influence the severity of drought.

Another hypothesis that drought climate associate EC is that the fumonisin B1 (FB1) contaminated corn may be involved in the carcinogenesis of EC, while drought climate and high temperature may influence the occurrence of fumonisins (Miller, 2001). The frequency of FB1 contamination in the EC high-risk areas is about twice that in the low-risk areas, and the average content of FB1 in samples from high-risk areas was approximately three times that from low-risk areas (Wang et al., 2000). As a result of the structural similarity between sphinganine and FB1, the mechanism of action of FB1 is mainly via disruption of sphingolipid metabolism, this is an important step in the cascade of events leading to altered cell growth, differentiation and cell injury. Toxicity of FB1 is given via inhibition of ceramide synthase that catalyzes the formation of dihydroceramide from sphingosine. This mechanism of action may explain the wide variety of health effects observed when this mycotoxin is ingested like high rate of human esophageal cancer (Soriano et al., 2005).

5. Conclusions

This study suggests an association between drought climate and mortality rate of esophageal cancer. The study also demonstrated that using existing health data, geographic and environment data, GIS and GIS-based spatial techniques could provide an opportunity to identify diseases within highly endemic areas, and also lay a foundation to pursue further investigation into the environmental factors responsible for disease risk. But the study was based on secondary data, which may have some limitations for epidemiologic studies. Environmental and climatic monitoring is needed to get more accurate data to determine the association between climate factors and esophageal cancer. In addition, other known esophageal cancer risk factors such as life style, biologic and genetic factors were not taken into account in this study, which may result in some ecological confoundings. There is a need for further studies that address the issue of potential confounding factors preferably using individual-level data for population and health.

Acknowledgments

We thank Dr. Guanghui Zhu and Dr. Xiangqun Ye for the helpful conversations and comments on the manuscript, Dr. Jianjun Zhang for the skilful technical assistance.

References

Akhtiamov MG, Kairakbaev MK. Incidence of esophageal cancer on the plains and in the mountainous regions of the Kazakh SSR [Russian]. Vopr Onkol 1983;29:49–53.