Geographic Information Systems (GIS) Application for Health: Case of Al Ain (UAE)

Yagoub, M. M.
Geography and urban Planning Department, College of Humanities and Social Sciences, United Arab Emirates University, P. O. Box: 17771, Al-Ain, E-mail: myagoub@uaeu.ac.ae, myagoub@hotmail.com

Abstract
Recently, there is an increase in the number of studies in UAE that examine the life style in relation to health and diseases. Example of these are the UAE Health and Life Style Survey 2000 (Badrinath et al., 2002) and the World Health Survey by WHO and Ministry of Health (Public Health, 2009). Although these studies try to provide a general trend of the common diseases and lifestyle in UAE, the majority of them did not put the geographic location and the environment surrounding it into consideration. The objective of this research is to assess the geographical distribution of health facilities and the impact of climatic conditions on diseases in Al Ain, UAE. Spatial and statistical analysis was used to aid in mapping the overall picture for health in Al Ain City, UAE. Geographic Information Systems (GIS) helped in linking the cases at clinics with population density, hence it provided a multi-layer analysis capability.GIS also provides a better modeling and visualization of the cases in spatio-temporal dimension. People who work in public and environmental health can get use of this study to demonstrate the applicability of GIS for analyzing health data.

1. Introduction

The use of geographical analysis for studying distribution of diseases is essential, this is because of the role of place in health (Jones and Moon, 1987). In the past and still now, medical geography represents the cornerstone for such analysis. Dr. John Snow (father of modern epidemiology) first linked the science of epidemiology with the use of geographic information to reveal relationships between environment and diseases (Longley et al., 2001). Approaches to geography of health take a variety of methodological perspectives, some researchers choosing to create statistical models of incidences or patterns of disease and environmental contamination (Openshaw et al., 1987), others adopting qualitative approaches to develop understanding of ill-health, or exploring the geographical expression of the politics of health (distribution, delivery, access to health services and population) (Hodgson, 1988 and Gatrell and Senior, 1999). Studies of disease distributions at local and regional scales commence with notification of disease among a set of individuals. If addresses of such individuals are available then the disease data can be mapped as a set of "point events". Alternatively, the cases may be aggregate to form a set of counts for a small area (e.g. census unit or post code) to avoid confidentiality issues (Marshall, 1991). Generally, the power of GIS is in managing, depicting, visualizing, analyzing, and communicating the health data in tabular, graphical, and map format. More details about the benefits of GIS in health can be found in Gatrell and Senior (1999), Lepper et al. (1995), Scholten and Lepper (1991). New areas of GIS application include its use in patient care (Boyer, 2002 and Hirschfield et al., 1993), online health query (Peters and Hall, 1999; ESRI, 2000), telemedicine and management of rooms (Glasgow, 2003), and development of specific GIS software for health (Geohealth, 2003). Within the broad context of GIS for health, this study is an attempt to assess the characteristics of the health in Al Ain City, UAE.

2. Objectives

This study is designed to address the followings:

1. To determine the geographical distribution of health facilities using spatial statistics. The null hypothesis assumed here is that health facilities in Al Ain are randomly distributed.
2. To evaluate spatial and space-time disease clusters. It is assumed that the industrial area at the Southern part of the City represents a source of pollution and there is a possible clustering of disease associated with it.
3. To study the impact of climatic conditions (temperature) on disease. Temperature is assumed to have positive correlation with the number of cases.

3. Methodology

In this study spatial statistical functions and correlation coefficients were used to aid in the analysis. The geographical distribution of health facilities was analyzed using the spatial statistical functions such as geometric center, the Nearest Neighbor Index (NNI), K-function, and quadrat analysis. The geometric center of a point distribution is represented by a point location specified by the mean of x and y coordinates. The NNI measures the spatial arrangement of point features (clustered, scattered, uniform, and random). The K-function describes the relative locations of pairs of point events at different distances. The K-function (Ripley, 1981) considers all inter-event distances rather than just those to nearest neighbors, K(d) being defined as the expected number of further events within a distance d of an arbitrary event. Quadrat analysis of point patterns requires overlaying quadrats (grid) onto a map of point features in order to examine the distribution based on frequency of occurrence rather than inter-separation distance. In this case, the evaluation of spatial patterns is determined by the frequency of points within each quadrat. The Correlation coefficient (r) was used in this study to investigate the impact of climatic conditions on diseases. The Correlation coefficient indicates the extent to which the value of another variable is related to other variable. The value of the correlation coefficient ranges between –1 and 1. A positive value indicates that there is a positive (direct) correlation between the two variables, that is, a larger value of one variable implies a larger value of the other. At the extreme, r = 1, the two variables are perfectly correlated and their distribution patterns must be identical. A negative r value implies a negative (indirect) correlation; the variables are inversely related to each other. A higher value of one variable implies a lower value of the other (Chou, 1997). If the r value is not significantly different from 0, there is no correlation between the variables and the distributions of the two variables are considered independent of each other. Attribute data about diseases at each clinic were keyed in Microsoft Access and linked to the geographical location of clinics under GIS software. The GIS was used as an environment from which all analysis related to the objectives outlined above were carried out. The geographical distribution of health facilities was analyzed using the spatial statistical functions. The spatial statistical functions were not originally integrated with the GIS software, but special Visual Basic for Application (VBA) codes were created for them. To study the geographical distribution of diseases in Al Ain, the City was divided into three regions and the percentages of cases at each region were compared with population density. For evaluation of the spatial and space-time disease clusters, it was assumed that the industrial area at the Southern part of the City represents a source of pollution and there was a possible clustering of disease associated with it. This was due to availability of workshops, factories, printing presses, squatter settlements, and low socio-economic conditions that resulted in a large number of men living in a comparatively limited space and unhealthy environment (limited sanitation, sewerage, drainage). The spatial variation in the Diabetes Mellitus (DM) and Hypertensive Disease (HD) from the industrial area was tested using the distance as the main indicator. The centroid of the industrial area was taken as an origin from which diseases were assumed to diffuse and the location of the clinics were taken as centroids for the diseases. The relationship between the distance from the industrial area and number of cases was modeled and its strength was tested using coefficient of determination (R^2). To study the impact of climatic conditions on disease (correlation between temperature variation and the number of cases), average temperature and cases were used. Though the study area contains substantial geographic variations, particularly variation in elevation environment, corresponding variation in temperature is less apparent. This is because of the small area of the highland in the City and the total area covered by the City as a whole (646.497 square kilometres). Therefore, the average temperature records between 1982 and 2002 at Al Ain International Airport were taken as representative to the City. The average temperature was taken as *independent* variable and the average cases for Diabetes Mellitus (DM) and Hypertensive Disease (HD) at all clinics were used as *dependent* variables

4. Scope of the Study: Study Area and Diseases

Al Ain City was selected as study case due to availability and accessibility to some geographical and health data. Al Ain is the largest City in the Eastern region of the UAE. It is located approximately between latitude 24 ° 03′ and 24 ° 22′ North and longitude 55° 28′ and 55° 53′ East (Figure 1). The population of Al Ain City according to the 2001 census was 260,175 people. With the increase in population and expansion of residential areas, the study of spatial distribution of health facilities and diseases are vital for the public, the Preventive Medicine Department, and other institutions that involved in epidemiology of the City. Two diseases for the years 2000, 2001, and 2002 were selected for this study. The diseases are Hypertensive (Blood

Pressure) and Diabetes Mellitus. The selection was based on the fact that they represent the highest rates in comparison to other diseases. It has to be pointed out that Hypertensive or diseases of the circular system in general are among the top causes of death in the City (Al Ain Preventive Medicine Department, 2000). The period 2000, 2001, and 2002 was selected because of the availability of data.

5. GIS Database Sources

The GIS database includes geographic and attribute data. Geographic data such as roads and district boundaries were obtained from the Town Planning Department in Al Ain. In addition to that, some other geographical features such as locations of clinics (healthcare centers) were collected from field survey using SILVA Global Positioning System (GPS) receiver. Land use data were extracted from the IKONOS satellite image of 2001. The land use data were used to identify the industrial areas. The image was also used as a background to aid in analysis. Attribute data for diseases registered at each clinic were obtained from the Head Quarters of the Clinics in Al Ain Hospital and census and population data were obtained from Abu Dhabi Planning Department. The climatic data (mean temperature) were obtained from Al Ain International Airport.

6. Results and Analysis

The analysis focused mainly on geographical distribution of health facilities using spatial statistics, geographical distribution of diseases, and the impact of climatic conditions on disease (correlation between temperature variation and the number of cases).

6.1 Geographical Distribution of Health Facilities using Spatial Statistics

The total number of governmental clinics inside Al Ain City is 11 (9 district clinics and 2 hospital clinics). The geometric center of the health facilities was found at an eculidean distance (direct) of 1,964 meters and functional distance (following the road) of 3,065 meters from the main hospital in the City. This indicates that the main hospital is located at a suitable proximity from the clinics. Nearest Neighbor Index (NNI) values relate how clustered or dispersed health facilities are. An NNI value of 0 (zero) indicates an intensely clustered pattern, while an NNI value of 1 indicates a random distribution, and an NNI value of 2 (or higher) indicates a strongly dispersed or organized pattern (Chou, 1997). To test the statistical significance of computed NNI, the standard normal deviation (z) was used. The computed z value was compared with the normal distribution value of 1.96 for level of significance ($\alpha = 0.5$), to test the hypothesis of whether the spatial pattern is random. The null hypothesis that health facilities in Al Ain are randomly distributed was accepted with NNI = 0.72 and $|z| = 1.75$. Odd distances of 3, 5, 7, and 9 kilometers were taken as examples for testing the number of health facilities within each distance. The distances give a general idea about how close the health facilities are to each other. Using the K-function, the average number of health facilities one expects to find within 3 kilometers (km) of another facility was estimated as 0.727. The average number of health facilities one expects to find within 5 km, 7 km, and 9 km of another facility were estimated as 2.36, 3.82, and 5.64, respectively. It was deduced that as the distance increases, the number of expected health facilities also increases. Table 1 shows an example of distance matrix for health facilities in Al Ain. The matrix provides a measure of how far the distance between the facilities. The distances were found range between 3125.498 meters and 16345.138 meters (Table 1). Generally, more than 70% of health facilities were found to be within a buffer distance of 5 km from each other and 91% within a buffer of 9 km from each other. The quadrat analysis is based on the null hypothesis that point features are randomly distributed in the Poisson distribution. Using quadrat analysis (the City was divided into 10x10 squares grid) upon the health facilities. The average number of health facilities per quadrat was found equal to 0.11 with a variance of 13.79 and variance to mean ratio of 125.36. Ninety one percent (91%) of the quadrats were found with no health facility, 7% with one health facility, and 2% with two health facilities.

Table 1: Example of distance (in meters) matrix between health facilities in Al Ain

Name of health facility	AlHeely	AlMasoudy	Town	Tawam Hospital	Al Ain Hospital
AlHeely	0.000	3750.637	7034.443	16345.138	7322.894
AlMasoudy	3750.637	0.000	5277.019	12597.555	4110.606
Town	7034.443	5277.019	0.000	12455.949	3125.498
Tawam Hospital	16345.138	12597.555	12455.949	0.000	9815.105
Al Ain Hospital	7322.894	4110.606	3125.498	9815.105	0.000

The result demonstrated a random distribution to the health facilities in the City and it was similar to the conclusion obtained from the Nearest Neighbor Index (NNI) analysis. Accessibility surfaces were constructed as a method of determining spatial inequalities among health facilities areas and districts across the City. The distribution revealed that the majority of the health facilities (85%) were located in attractive areas along major roads and in proximity to population. However, it was found that 76% of the private health facilities were concentrated at the City Center. This is due to the high return from facilities located near the City Center (more visitors). Some districts with 60% of the total population in the City have two health facilities within one to four kilometers while others (40% of the population) are at a distance of more than nine kilometers. This implies that the geographical distribution of health facilities is not even and more health facilities are needed, especially in suburban districts.

6.2 Geographical Distribution of Diseases

Clinics were divided into three main regions based on their directional location from the weighted mean center of all the clinics (Table 2). Based on this division, Diabetes Mellitus (DM) was found more prevalent in the Northern, Southern, and Southwestern direction of the city (Table 2). The distribution of DM was correlated with population income and a strong relationship was found with class of population i.e. areas dominated by people of high income have more cases than other areas. High percentages (35% and 39%) of Hypertensive Disease (HD) cases were found near the center of the city and in the Northern part of the city (Table 2). Strong relationship between higher percentage of cases and population density of districts was found (Figure 1).

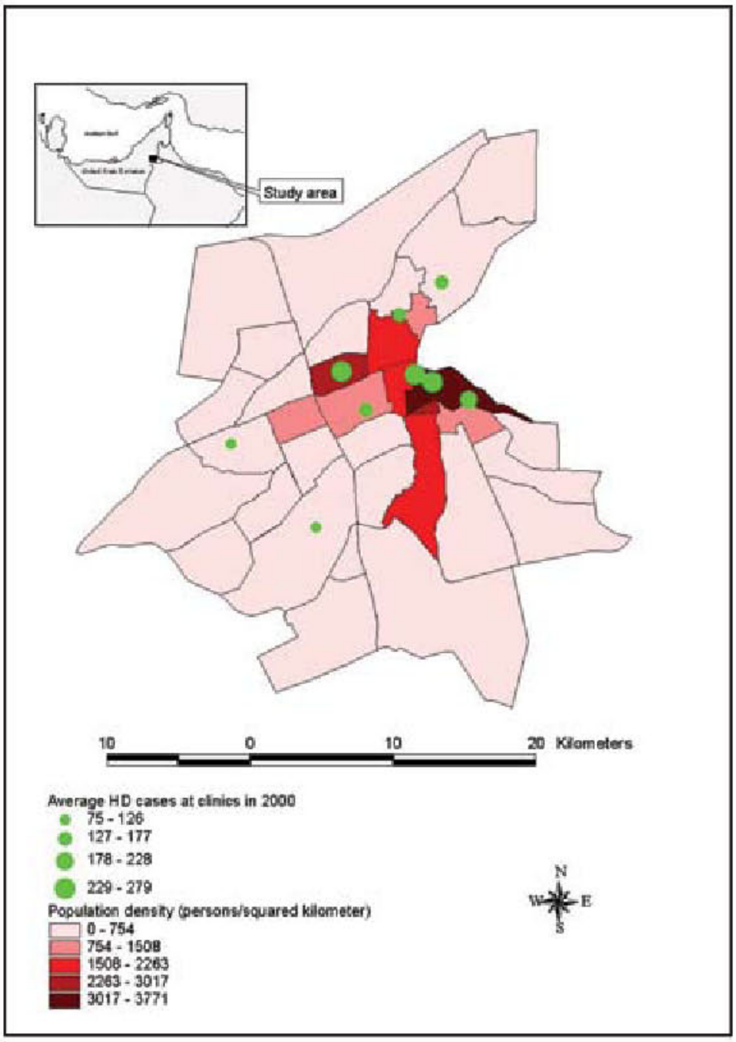

Figure 1: Hypertensive Disease (HD) cases at Al Ain in 2000 in comparison to population density

Table 2: Geographical distribution of Diabetes Mellitus (DM) and Hypertensive Disease (HD) between 2000 and 2002

Region	Number of clinics	DM cases	%	HD cases	%
City Center (East)	3	1264	30.1	1487	35.1
North	3	1566	37.2	1641	38.7
South and Southwest	3	1376	32.7	1109	26.2
Total	9	4206	100	4237	100

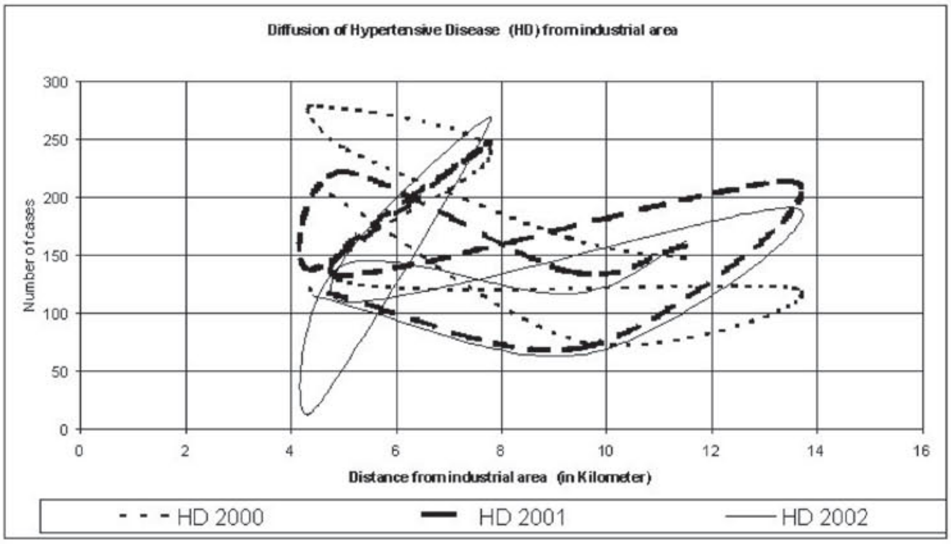

Figure 2: Diffusion of Hypertensive Disease (HD) with space and time

Therefore, its relation to the distribution of population can indicate clustering of HD in the City. Little significant spatial variation (clustering) in Diabetes Mellitus (DM) and Hypertensive Disease (HD) from the industrial area was discernable. The strength of the relationship between the distance from the industrial area and number of cases was evaluated by using coefficient of determination (R^2). The coefficient measures the proportion of variation in the number of cases accounted for by distance from the industrial area. The relationship was weak, because the coefficient of determination (R^2) for the 6 linear models was low (ranging between 0.0211 and 0.4914). This means that only 2.11 to 49.14 per cent of the variation in the number of cases can be accounted for by a linear function of the change in the distance from the industrial area. 66% of the models proved that as the distance from the industrial area increases, the cases also increase and this indicates that the industrial area does not represent a point source of diffusion for Diabetes Mellitus (DM) and Hypertensive Disease (HD) (Figure 2). Serial data of the diseases at geographic areas (clinics) were displayed as a scatter plot of cases versus distance from the industrial area. All clinics were found located at distances between 4 kilometers and 14 km from the industrial area and this was why Figure 2 appeared to be shifted by 4 km. By presenting cases for three years (2000, 2001, 2002) on the same axis, the scatter plot permits direct comparison and gives a spatial-temporal summary of the diffusion process in relation to the industrial area (Figure 2). The diffusion waves of both Diabetes Mellitus (DM) and Hypertensive Disease (HD) had no foci and no indication of propagation from the industrial area (Figure 2). Therefore, the industrial area cannot be considered as a point source of pollution for these diseases. However, the process is not this simple, and hierarchical diffusion between major highly populated districts undoubtly contributes. Diseases are never going to be uniformly distributed across a geographic region. Some clustering is inevitable and can be explained away as due purely to chance. Local residents generally do not like these explanations because many communities live near potentially hazardous land uses (e.g. industry) feel that such explanations are too dismissive of their health fears. Geographical distribution of diseases analysis depends on the accuracy of the data provided by the Head Quarters of the Clinics in Al Ain.

There is a possibility of errors encountered in diagnosis of cases and in the statistics provided. Moreover, people who live in certain areas during a certain period of time may shift their locations. This is especially true in UAE, where a great in and out migration among expatriates is noticed, putting into account that expatiates represent almost 70% of the total population. Such frequent migration may change the map of the geographical distribution of diseases from time to time.

6.3 The Impact of Climatic Conditions on Disease (correlation between temperatures and the number of cases)

Climatic conditions such as temperature have impacts on human health (Kalkstein and Greene, 1997 and McMichael et al., 1996). For example, high temperatures have significant adverse effects on mortality and morbidity. Rooney et al., (1998) estimated excess mortality in England and Wales associated with the 1995 heat wave at 619 deaths (8.9% increase). The objective here is to check whether there is any correlation between temperature as independent variable and number of cases for Diabetes Mellitus (DM) and Hypertensive Disease (HD) as dependent variables in Al Ain City. Analysis showed that DM in 2000 and 2002 had low positive correlation coefficient with temperature, 0.346 and 0.493, respectively (Table 3). However, in 2001, DM showed a large negative correlation coefficient (-0.625). This implies that the DM is inversely proportional to temperature (Table 3). The correlation of HD in both 2000 and 2002 with temperature was positively low (0.361 and 0.247) and in 2001 it was negatively low (-0.397) (Table 3 and Figure 3). Generally, the correlations for both DM and HD with temperature were varying between positive and negative, therefore, it was concluded that the number of cases for both DM and HD in Al Ain City were independent on the temperature. For a better understanding, the relationship between number of cases and temperature for both DM and HD was modeled with a linear form.

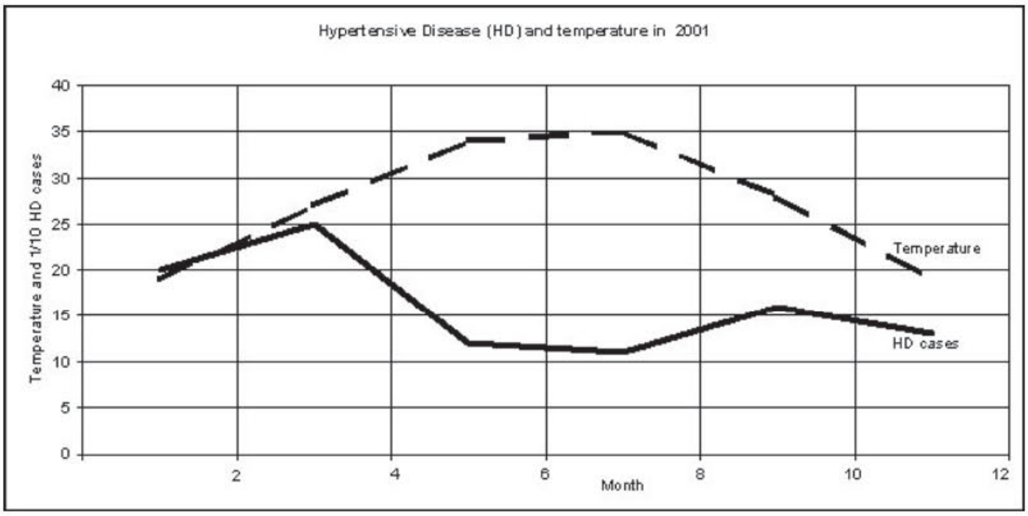

Figure 3: Relationship between Hypertensive Disease (HD) and temperature in 2001

Table 3: Correlation between temperature and the number of cases for Diabetes Mellitus (DM) and Hypertensive Disease (HD) between 2000 and 2002

Month	Temperature*	DM2000	DM2001	DM2002	HD2000	HD2001	HD2002
2	19	206	190	109	174	198	120
4	27	214	194	110	209	245	123
6	34	189	117	317	176	119	147
8	35	182	106	103	172	112	122
10	28	199	126	98	205	157	142
12	19	109	137	96	136	128	136
Correlation coefficient		0.346	-0.625	0.493	0.361	-0.397	0.247

* Average temperature records between 1982 and 2002

The strength of the relationship between the number of cases and temperature was evaluated by using coefficient of determination (R^2). The coefficient measures the proportion of variation in the number of cases accounted for by change in temperature. The relationship was weak, because the coefficient of determination (R^2) for the 6 models was low (ranging between 0.0609 and 0.3911). This means that only 6.09 to 39.11 per cent of the variation in the number of cases can be accounted for by a linear function of the change in temperature. Therefore, there was no strong causal relationship between the number of cases and temperature. This may be due to the spontaneous adjustments of population to the changed climate by improvements in housing (air conditioning), clothing, and behavior. However, there was a noticeable decrease in the number of the DM and HD cases during the hot season (July/August) (Figure 3). This is may be attributed to the summer vacation when a large number of people leave the City. For better understanding of the effect of climate on health, detailed mortality, morbidity, and socio-economic data are needed.

7. Conclusions and Recommendations

There have been considerable advances in the development of methods for the detection of clustering of health facilities and events, together with productive links between statistician, epidemiologists, and geographers in demonstrating the usefulness of GIS-based approaches. Using GIS delivers a wider variety of functionality through its customization capabilities, enabling users to make more advanced spatial queries and modeling. In this study, GIS was used to analyze health data in Al Ain. For example, the geographical distribution of health facilities were analyzed using spatial statistical functions. The main hospital is found in a suitable location for the majority of the residents in AL Ain. The health facilities were found randomly distributed with Nearest Neighbor Index (NNI) of 0.72 tested at standard normal deviation of 1.96 at 0.5 level of significance. Little significant spatial variation in the Diabetes Mellitus (DM) and Hypertensive Disease (HD) from the industrial area was discernable. Generally, the correlations for both DM and HD with temperature were varying between positive and negative, therefore, it was concluded that the number of cases for both DM and HD in Al Ain City were independent on the temperature. Output in forms of maps, scatter plots, models, and reports showing rates and geographical distribution of diseases were generated. The output can be used by various health sectors in Al Ain to help in identifying needs and allocating resources. Other environmental data (possible point source of pollution) that will be added to the database includes location of dumping sites, electricity lines, telecommunication and TV receiving stations. Future work will investigate the association between diseases with sources of pollution. GIS database for roads and health facilities together with centroids of districts will be used to determine optimum routes for patients and service areas. Top causes of death diseases and those with higher rates such as malignant neoplasm, acute rheumatic fever, leukemia, chickenpox, mumps, pulmonary, and tuberculosis will be geographically analyzed. The work can finally be put in a form of a health atlas and deployed on the Internet for public access. Low cost hardware, GIS software, and availability of spatial data in digital format are putting GIS within the grasp of smaller health departments. However, the big block is the availability of health data (patients records) in digital format. Therefore, it is recommended that registration of patients' details and cases to be done using digital database. The issue of confidentiality with regards to geo-coded health data on individuals can be overcome by blocking (masking) all individual particulars (name, telephone, etc.) and limiting the search for age, gender, the cases, their date, and geographical locations (district name and community number). Moreover, cases can be given as aggregated values for a district or a community; therefore, there is no reference to particular house or person.

Acknowledgements

The author would like to acknowledge the cooperation of the Town Planning Department, Preventive Medicine Department, and Head Quarters of the Clinics in Al Ain. The comments and suggestions made by Professor G. G. Garland, former Chair, Geography Department, UAE University are highly appreciated. Credits goes to the reviewing team and editing by the International Journal of Geoinformatics.

References

Al Ain Preventive Medicine Department, 2000. Annual Statistical Report 2000. Al Ain Preventive Medicine Department.

Badrinath, P., Al-shboul, Q., Zoubeidi, T., Gargoum, A., Ghubas, R., and El-Rufaie, O., 2002, Measuring the Health of the Nation: *UAE Health and Life Style Survey 2000 (UAEHALS2000)*. Published by the College of Medicine and Health Sciences and College of Business and Economics, Sponsored by the Research Affairs, UAE University, Al Ain.

Boyer, G., 2002, Downey Regional Medical Center Provides High-Quality Patient Care through GIS. *ESRI ArcNews,* 24 (3), 30.

Chou, Y. H., 1997, *Exploring Spatial Analysis in Geographic Information Systems*, (Santa Fe: OnWord Press).

ESRI, (Environmental Systems Research Institute), 2000, Web Sites using ESRI IMS Software: from the Baltic Sea to the Twin Cities, Internet Mapping Servers the world. *ESRI ArcNews,* 22 (4), 19.

Gatrell, A. C., and Senior, M., 1999, Health and Health Care Applications. In: Longley, P. A., Goodchild M. F., Maguire D. J., and Rihnd D. W. (eds). *Geographical information systems: Management issues and applications*. Vol.2: 925-938, (New York: John Wiley and Sons).

Geohealth, 2003, BodyViewer. http://www.geohealth.com/bodyviewer.html (accessed on 22/9/2003).

Glasgow, A. B., 2003, Telemedicine in the Gulf. In: Nancy Sappington, *Geography and GIS serving our world, ESRI Map Book,* volume 18: Redland: ESRI Press: 73.

Hirschfield, B., Brown P., and Bundred P., 1993. Doctors, Patients, and GIS. *Mapping Awareness,* 7 (9), 9-12.

Hodgson, M. J., 1988. An Hierarchical Location-Allocation Model for Primary Health Care Delivery in a Developing Area. *Social Science and Medicine,* 26, 153-61.

Jones, K., and Moon, G., 1987. *Health, Disease and Society: An Introduction to Medical Geography*, (New York: Routledge & Kegan Paul Books Ltd.).

Kalkstein, L. S., and Greene, J. S., 1997, An Evaluation of Climate/Mortality Relationships in Large US Cities and the Possible Impacts of a Climate Change. *Environ Health Perspect,* 105, 84-93.

Lepper, M. J., Scholten, H. J, and Stern, R. M., 1995, *The Added Value of Geographical Information Systems in Public and Environmental Health* (Dordrecht: Klwer).

Longley, P. A., Goodchild, M. F., Maguire, D. J., and Rihnd, D. W., 2001, *Geographic Information Systems and Science,* (New York: John Wiley and Sons, Inc.).

Marshall, R. J., 1991, A Review of Methods for the Statistical Analysis of Spatial Patterns of Disease. *Journal of the Royal Statistical Society,* Series A, 154, 421-444.

McMichael, A. J., Haines, A., Slooff, R., and Kovats, S., 1996. *Climate Change and Human Health: an Assessment Prepared by a Task Group on behalf of the World Health Organization*, the World Meteorological Organization and the United Nations Environment Programme, (Geneva: World Health Organization).

Openshaw, S., Charlton, M., Wymer, C., and Craft A., 1987. A Mark 1 Geographical Analysis Machine for the Automated Analysis of Point Data Sets. *International Journal of Geographical Information Systems* 1: 335-558.

Peters, J. and Hall, G., 1999. Assessment of Ambulance Performance using Geographic Information Systems. *Social Science and Medicine,* 49, 1551-1566.

PublicHealth, 2009, World Health Survey. (http://www.publichealthsurvey.com, accessed on Jan. 2009)

Ripley, B. D., 1981. *Spatial statistics*, (New York: John Wiley and Sons).

Rooney, C., McMichael, A. J., Kovats R. S., and Coleman, M. P., 1998. Excess Mortality in Greater London, during the 1995 heatwave. *Journal of Epidemiology Community Health,* 52, 482-486.

Scholten, H. J., and Lepper M. J. C., 1991, The Benefits of the Application of Geographical Information Systems In Public And Environmental Health. *World Health Statistics Quarterly,* 44, 160-170.