REVIEW

Geographic Information Systems (GIS), an Informative Start for Challenging Process of Etiologic Investigation of Diseases and Public Health Policy Making

Alireza Mousavi Jarrahi[1], M Zare[2], A Sadeghi[3]

[1]Department of Social Medicine, Medical School, Shahid Beheshti University of Medical Sciences and the Cancer Institute Research Center, Tehran University of Medical Sciences Tehran, Iran. [2]Ministry of Health, Deputy for Research, Tehran, Iran. [3]Department of Remote Sensing and GIS, Shahid Beheshti University, Tehran, Iran.

Abstract

Background: The public health has been always concerned of the immediate environment of human as causal factors for different diseases and health outcomes. Epidemiology, as one of the fundamental basis of public health, is concerned of how diseases are distributed in terms of geographical, chronological, and human population characteristics and employees the descriptive nature of such spread to draw conclusion on the etiology of health or disease utcome for further policy making on prevention of disease or promotion of health. **Methods:** In this paper, we present the importance of GIS technology in epidemiology from both descriptive and etiologic standpoints and elaborate how this technology can stand in the forefront of disease and health outcome measures in the coming decades. The paper will address the history of geo-related health and disease issues. The mapping tool as a traditionally strong resource in the public health will be explored. The advances in Information Technology and one of its best utilized offshoot, GIS, in Health and disease will be discussed. How the huge repository of generated or ever generating geo-related data and information is utilized to address etiology of diseases or to help public health authorities in making informed policy making decisions are explored. **Results:** The utilization of GIS technology in diseases with intermittent host such as malaria, yellow fever, or other parasitic diseases has already been well established. The GIS technology and its utilization in chronic and degenerative diseases such as cancer, diabetes, and aging are under development and new frontiers are discovering. The limitation of GIS technology in addressing host environment interaction in micro-environment (at the molecular biology and tissue pathogenicity level) and gene–environment interaction (at the individual level) will further be discussed. **Conclusion:** We then distress on the efficient use of GIS both in the etiologic investigation of diseases and health events as well as the utilization of the GIS technology as a administrative tool in the help of public health authorities and policy makers in strategic management of health of a community or emergency management of man-made or technological disasters (e.g., wars) or naturally occurring disasters (e.g., earthquake and floods).

Keywords: Epidemiology- GIS- Iran- public health

Introduction

In order to address the diverse features of disease outbreakings, scientists have continually been seeking out different aspects of an epidemic from cell proliferation at the microenvironment of the disease pathogenicity to large-scale diversities and variations seen in the incidence of disease and prevalence of exposure to potential or known hazards across different geographic boundaries. Notwithstanding striking advances in cellular and molecular technology dealing with disease pathogenicity has occurred, the notion of discoveries of new frontiers in fighting diseases relays on the development of new thoughts, assumptions and,

Corresponding Author:
Dr. Alireza Mousavi Jarrahi
Department of Social Medicine, Medical School, Shahid Beheshti University of Medical Sciences and the Cancer Institute Research Center, Tehran University of Medical Sciences Tehran, Iran.
Email: rmosavi@yahoo.com

in scientific words, hypotheses. The starting point of scientific discovery is the formation of a hypothesis generated by the observation of a series of events and heuristically examination of these events against the established body of scientific facts in order to advance the frontier of discerning a biological or physiological phenomenon.

Epidemiology, as the science of population and health study at the macro-environment level, develops and challenges hypotheses for all aspects of public health concerns include disease prevention and health promotion. Descriptive epidemiology,the portrayal of disease or any health event at the three dimensions of place, time and human population, has been the pillar of hypothesis formation ever since John Snow hypothesized that cholera is a waterborne disease with employing map as a tools to visualize the cholera distribution across urban geographic scale [1].

In this paper, we portray the importance of GIS technology in epidemiology from both descriptive and etiologic prospects and illustrate how this technology can stand in the forefront of disease and health outcome measures. Geography and Health Place, in general term or Geography in real life term as one of the triad of descriptive epidemiology (place, time, Human biology), is used to indicate a mixture of lifestyle, physical environment, and genetic factors (common genetic pool). While there is no exact definition of what a place is, the political boundaries at the international level and the administrative boundaries at the national level make the core practicalities of a place where disease variation can be measured or described. The disease incidence is mainly measured on the administrative boundaries where population counts are available. The purpose of any study exploring the variation of disease incidences across geographical definition or boundaries (or, as it is loosely called, place) is to identify the possible causes in order to explain the variation in disease occurrence for further understanding of disease etiology or prevention related policy making strategies. In public health, the study of geographic ariation has a broad unitization and they are done in different scales, from variation seen at the continent level, to variation seen in urban areas and clusters of diseases cases around some sources of hazards [2-5].

The introduction of information technology and utilization of geographic information systems (GIS) have provided valuable opportunities in describing these geographic variations and furthering the development of hypotheses regarding the etiology of diseases [6]. Geographic variation in cancer (as a disease of modern era) occurrence is seen in almost all geographic scales. There are distinct patterns of site-specific cancer occurrence in the developed versus the developing countries.

Cancer in certain sites, such as the stomach and esophagus [51], are generally more common in the low-resource countries in Asia (except for Japan, which has a high incidence of stomach cancer although it is considered a developed country), cervix in Latin America and the Indian subcontinent and cancers of breast, prostate and colon are more common in the affluent countries of North America and Europe [7, 8]. Along with the variation seen on scales as large as continents and countries, there is variation in cancer occurrence within a country and sometimes in distances as small as 100 kilometers. Such small places with distinct patterns of cancer, termed as hot spots, are seen in Europe, exemplified in esophageal cancer in France or Iran, or other parts of the world [9]. There are very diverse methodologies to study the variation of any disease in different places. Most of these methodologies are based on the incidence and mortality rates, where incidence is defined using some arbitrary boundaries or census tracts, depending on the availability of information about the population or enominator. New techniques, such as GIS and remote sensing technologies with complex analytical procedures have been used to draw hypotheses on the etiology of different diseases, including cancer across geographic scales [10]. The tools in hands of geographer has brought promising prospect for complicated techniques of disease cluster detection as well as addressing geographic variations at smaller scales.

The sources of data available to address geographic variation of diseases (cancer as example) To address geographical aspect of health and diseases, one needs the information and data as well as the tools to interpreted information or to analyze the data. While each disease or event of health yields its own repository or sources of geo related data, available data to ancer epidemiologist will be presented hereto address how the new tools such as GIStechnology can be utilized in the services of public health. In addition to several ad hoc resources available to the cancer epidemiologist to address geographic ariations, there are two important sources with a vast repository of data. The most useful resource is the Cancer Incidence in Five Continents (CI5), published by the International Agency for Research on Cancer.

Currently, with eight volumes published thus far, the CI5 reports the data of cancer registries from around the world that have met the high standards of completeness and validity. Since the CI5 publication is centrally managed, its published data are comparable, making it a very valuable resource for epidemiologists interested in geographical variation in cancer incidence across the globe. The CI5 provides incidence data for more than 30 site-specific cancers by age group and sex [11] .

The second main source of this type of cancer information is the WHO mortality database. This database also has the benefit of comparability in terms of the coding system used, because mortality around the world is reported using different ICD (International Classification of Diseases) coding versions. However, mortality data is always subject to issues of validity at the point of establishment of cause of death in the death certificate. The practice of establishing cause of death varies from country to country, depending on autopsy rate and other influencing factors. In addition, the fact that mortality data is highly dependent on the quality of health care delivery and screening programs, interpretation of the mortality data across the different geographical regions could be misleading, especially when it comes to drawing conclusions on the etiologic aspect of cancer.

Tools available to explore geographical variation of disease, GIS and disease mapping Maps have long been a tool to describe the distribution of disease based on geographical scales. While the very first epidemiologic investigation of cholera by Jon Snow is based on the geographic distribution of mortality from cholera, the first maps or atlases of disease frequency were published in 1930, describing the geographic variation in cancer mortality in England and Wales [12]. A survey of resources in disease mapping in 1991 revealed approximately 49 international, national, and regional disease atlases [13].

Maps convey instant visual information on the spatial distribution of diseases and can identify subtle patterns, which may be missed in tabular presentation. They usually portray variation in the occurrence of disease morbidity or mortality mainly related to underlying sociodemographic indicators. Nonetheless, the map is the basic tool of visualization used to formulate hypotheses ranging from the disease etiology to the effect of service delivery, survival as well as mortality [14].

While disease mapping has the benefit of visualizing differences in the distribution of diseases across a given area of interest, the degree of differentiation is based on visual perception, which is influenced by various features of the map such as the plotting symbols used (types of shading, simple boundary maps, and so forth). Considerable caution is required regarding to avoid over-interpretation when dealing with disease maps. An empirical study has found that the way a map presents data has much more effect on observer perception of the spatial variation as would the actual differences existing in the data [15]. Another important issue in mapping disease occurrence is the choice of summary measure presented and visualized in the maps. In any case, since maps and visualization carry the main notion of comparing some measure of disease frequency, caution must be taken regarding the choice of summary measure, which two categories are available for mapping disease occurrence. The first is descriptive measures of frequency, incidence, mortality, prevalence, as well as relative frequency, and the second includes measures of associations such as standard incidence ratio, standard mortality ratios, proportionate mortality and morbidity ratios. When a measure of association is used to map disease data, each segment of the map is presented as a ratio of a standardized rate in that segment to a reference rate in a population or place, and the choice of reference population has to be carefully considered, especially when the variation across the mapping area is not very large. Geographic information systems (GIS) capture, store, retrieve, analyze and display spatial data in an automated manner [16]. GIS treats spatial data as unique since it is linked to a geographic map. The main components of GIS, in addition to a database, include spatial or map information and tools that link and perform spatial analyses of disease or any health-related event. The GIS provide unique tools to perform three general types of spatial analysis tasks: visualization, exploratory data analysis, and model building [17].

Used in new ways to explore data from traditional statistical analyses, visualization shows variation in values over an area by the locations of outlier and influential values on maps, thereby enhancing epidemiologic research. Although such tools are being developed and explored, they would benefit greatly from a closer and more seamless link between the statistical packages and GIS. Visualization has been a basic tool of presenting and summarizing data either through a simple graph or complex and overlying maps. Exploratory spatial analysis, as the second class of GIS methods, lets the analyst intelligently search the spatial data, identify spatial patterns of interest, especially spatial clusters of disease, and prepare hypotheses to direct research. The third class of spatial analysis, modeling, uses procedures that test hypotheses regarding the etiology and transmission of disease.

By integrating standard statistical and epidemiologic techniques, the GIS produce data to use in epidemiologic models. Statistical analysis results are shown and modeling processes can then be displayed over space.

GIS technology and geospatial tools have been used at the United States' National Cancer Institute to: " 1) identify and display the geographic patterns of cancer incidence and mortality rates in the US and their change over time, 2) create complex databases to study cancer screening, diagnosis and survival at the community level, 3) assess environmental exposure using satellite imagery, 4) model spatial statistics in order to estimate cancer incidence, prevalence and survival for each US state, 5) communicate local cancer information to public health professionals and the public at large using interactive web-based tools, 6) identify health disparities at the local level by comparing cancer results across demographic subgroups, and 7) develop new methods of presenting geospatial data to communicate unambiguously to the public and to allow researchers to examine complex multivariate data".

GIS technology has opened a new window to the art of exploratory data analysis and visualizing the differences in disease frequency rates across large geographic areas. However, very often the building blocks of maps and the units of analysis in the GIS are the polygons comprising certain geographic or administrative boundaries. The measures of disease frequency rely on the incidence of disease in a defined population, which in turn depends on the boundaries drawn for administrative purpose, giving rise to enormous limitations in spatial modeling analysis which will be discuss in the following section.

Interpreting geographical variations Geographical variation seen in diseases occurrence is dependent upon the geographical scale in which the occurrence is observed. For example, interpretation of the observed variation in the geographical distribution of cancer frequency depends on three main components, including the geographical scale, the summary measures used in the analysis, and the technology and statistical complexity employed.

The geographical scale is an issue that brings precision to the interpretation of the observed variation. The differences in the incidence and pattern of cancer

occurrence at the continental scale has already established the frequency of cancers at different socioeconomic scales; the cancers of lung, colon, breast, endometrium and ovary are frequent cancers among relatively prosperous Europeans and North Americans and the cancers of esophagus, cervix, and stomach are the cancers seen mostly in the relatively poor countries [6]. The continental scale is so large that it includes a socioeconomic gradient with certain cancer frequencies across the globe, the interpretation of which is too broad and ambiguous and has very low value from the aspects of cancer control and etiology. Large scale geographical variation must be interpreted cautiously as such variation could be attributed to a range of factors, such as the differences seen in the patterns of cancer and magnitudes of occurrence between European and African countries that may be attributed to race, ethnicity, socioeconomic status and lifestyle as well as environmental hazards. Further classification of rates at this large scale may render it impossible to draw conclusions or generate hypotheses and is subject to difficulty and ambiguity. While it will be very hard to define a very large or very small geographical scale, the smaller the geographical scale the more refined the assessment of the etiologic attribute. For example, variations seen in the incidence of esophageal cancer in Linxian (China) and Gonbad (Iran) [18]has resulted in the generation of several hypotheses that have been tested and evaluated in order to understand the etiology of esophageal cancer.

Small-scale geographical entities, which include more homogeneous populations in terms of ethnicity, culture, lifestyle and nutritional habits, make evaluation and assessment of the etiology behind a variation less subject to ambiguity and ecological fallacy.

Another important factor in interpreting geographical variation is the summary measures that are used to describe or scale the magnitude of variations. As it is wellestablished, the incidence measures, when adjusted for influencing factors such as age and gender distribution, are the best measures for assessing geographical variations. Other measures, such as prevalence, PMR and MOR, are subject to certain comparability problems that may distort interpretation and terminate to wrong conclusions. The last factor in the interpretation of geographical variation studies is the use of the new technologies, statistical modeling and complexities to draw conclusions. One of the techniques used in the study of cancer, especially childhood cancers, has been the investigation of possible case clustering around certain environmental hazards such as landfills or nuclear reactors. The investigation of a cluster is, in fact, another way of describing smallscale geographical variation. This depends on a statistical model that incorporates a great deal of assumption. While the detection of cancer clusters are very important to epidemiologists looking at the etiology of cancer, detected clusters must be examined carefully, and the result of certain clusters requires other means of investigation to address the etiology behind the cluster [19].

The geography and health are two interrelated entities which the former complement the latter. The GIS, as a new technology, will provide ample opportunities for epidemiologists to understand the underlying factors affecting the health of people and for policy makers to draw better policy making strategies at the community level. This utilization of GIS is, in fact, an informative origin for challenging process of etiologic investigation and policy making of disease and health outcomes without misuse or over utilization of the technology.
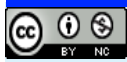
Acknowledgements

References

1. Johnson SB. The Ghost Map: The Story of London,s Most Terrifying Epidemic and How it Changed Science, Cities and the Modern World.2006.

2. Wu K LK. Association between esophageal cancer and drought in China by using Geographic Information System. Environ Int. 2007;33(5):603-8.

3. Garb JL, Ganai S, Skinner R, Boyd CS, Wait RB. Using GIS for spatial analysis of rectal lesions in the human body. International journal of health geographics. 2007;6:11.

4. Bunnell JE TC, Bushon RN, Stoeckel DM, Brady AM, Beck M, Lerch HE, McGee B,Hanson BC, Shi R, Orem WH. Possible linkages between lignite aquifers, pathogenic microbes, and renal pelvic cancer in northwestern Louisiana, USA. Environ Geochem Health 2006;28:577-87.

5. Thompson JA CS, Zhu L. An evaluation of spatial and multivariate covariance among childhood cancer histotypes in Texas. . Cancer Causes Control 2007;18:105-13.

6. Briggs DJ, Elliott P. The use of geographical information systems in studies on environment and health. World health statistics quarterly Rapport trimestriel de statistiques sanitaires mondiales. 1995;48(2):85-94.

7. Nasca C.Philip PH. <Fundamentals of Cancer Epidemiology.> (Aspen Publishers, Inc: Maryland USA)2001.

8. Muir C.S. NJ. International Patterns of Cancer. In: <Cancer Epidemiology and Prevention>, Eds: Schottenfeld D. and fraumeni J.F., Oxford University Press: Oxford, New York. 1996:141-67.

9. Munoz N. DNE. Esophageal Cancer. In: <Cancer Epidemiology and Prevention>. Eds: Schottenfeld D. and fraumeni J.F., Oxford University Press: Oxford, New York1996. 681-5 p.

10. Beck L.R. KU, Bobo M. . Remote sensing, GIS, and spatial statistics: Powerful tools for landscape epidemiology. In: <Health Impacts of Global Environmental Change: Concepts and Methods>. Eds: Martens P, Cambridge University Press. 2002.

11. Parkin DM WS, Ferlay J, Teppo L, and Thomas DB. . Cancer incidence in five continents. [VIII]. Lyon, France, IARC Press. IARC Scientific Publications No. 155 .2002.

12. Swerdlow A. D-SSI. <Atlas of Cancer Incidence in England and Wales, Oxford University Press: Oxford1993. 1968-85 p.

13. Walter SD, Birnie SE. Mapping mortality and morbidity patterns: an international comparison. International journal

of epidemiology. 1991;20(3):678-89.

14. Jarup L. Health and environment information systems for exposure and disease mapping, and risk assessment. Environmental health perspectives. 2004;112(9):995-7.

15. Walter SD. Visual and statistical assessment of spatial clustering in mapped data. Statistics in medicine. 1993;12(14):1275-91.

16. Clarke KC, McLafferty SL, Tempalski BJ. On epidemiology and geographic information systems: a review and discussion of future directions. Emerging infectious diseases. 1996;2(2):85-92.

17. Gatrell AC, Bailey TC. Interactive spatial data analysis in medical geography. Social science & medicine (1982). 1996;42(6):843-55.

18. Mosavi-Jarrahi A, Mohagheghi MA. Epidemiology of esophageal cancer in the high-risk population of iran. Asian Pacific journal of cancer prevention : APJCP. 2006;7(3):375-80.

19. De Stefani E, Parkin DM, Khlat M, Vassallo A, Abella M. Cancer in migrants to Uruguay. International journal of cancer. 1990;46(2):233-7.