Epidemiology is a branch of medical science that deals with the incidence, distribution, and control of disease in a population according to the Merriam-Webster Dictionary. This definition is important because it helps to define how GIS may be applicable to the study of disease. GIS can easily be used to map occurrences of disease, which leads to an understanding of the distribution of disease as the data set expands. These are key concepts to starting a control program to manage the disease. However, when starting to build a GIS for epidemiology the key elements needed are not quite as simple as stated above. This is why epidemiologists are starting to take a landscape ecology approach that will account for the multiple populations of hosts, vectors and pathogens, as well as the interaction of all of these populations and the environment (Kitron, 1998). It is this matrix of elements that makes GIS an attractive tool for spatial analysis of diseases.

From the articles reviewed, it was clear that GIS can be used in many ways to assist in epidemiological research. The uses range from putting break out locations on a map from GPS points to advanced modeling of population dynamics of a specific animal. There were several data models used. For example, a logistic regression model was used to determine risk factors for Lyme disease (Vine et al., 1997), while a model named Geographic Resource Analysis Support System (GRASS) was used to examine the population dynamics of red and grey squirrels (Rushton et al., 2000). In another study by Guerra et al. to determine habitat suitability of a tick species, *Ixodes scapularis*, all of the data used was coded and then a positive or negative analysis was used, followed by a logistic regression analysis using the environmental factors as independent variables and the positive and negative sites as outcome variables. Once this was done, the area of interest in Wisconsin and Illinois was turned into a grid with 2.5 km² cells and values were assigned to each cell using the Summarize Zones feature of ArcView. In the final step, the logistic equation was used to generate the probability of the presence of *Ixodes Scapularis*, which was used to create a habitat suitability map that will benefit Lyme disease management in the North Central United States (2002).

It is important to note that the predictive value of models is improved when the data used are accurate, the conditions for using the data are simple, and the area in question is close to the source of exposure (Vine et al., 1997). It was also emphasized several times in the articles reviewed (Vine et al., 1997; Mott et al., 1995; Barnes et al., 1999; and Hendrickx et al., 2001) that the use of GIS reaches beyond the technical capabilities for which it is known. GIS based projects also function as a reason to start communications with other agencies and ultimately enrich the management of a disease and ecosystem, simply by opening the gates of communication. The raw data will in the end
Effect management of diseases, even if it is just because everyone decided, for example, that less insecticide and variation in growing seasons is a better management plan for the community than intensive use of insecticides that does not really get the job done in the first place and coincidentally is bad for the ecosystem (Barnes et al., 1999).

Now that it is clear that GIS is relevant to epidemiology, a sample of the analytical tools within ArcInfo and ArcView will be presented. For example, when it is important to link people with a particular characteristic in the analysis, address matching is utilized. Whereas, if the user wants to visualize the distance between two points, then the distance function should be implemented. For example, the distance between the closest mosquito breeding water hole and the community affected. Buffering is another good tool when considering control regimes, because it will make it possible to see how far an area of ten miles is from the mosquito breeding area. If a user wanted to know the areas affected or the locations of habitat most suitable for the animal carrying the disease, then spatial query is the best tool to use. Polygon overlay analyses are also important, since this is the best way to integrate the attributes of two data layers into one. The end result of one kind of overlay will be the areas where both polygons overlap. For example, where rivers in forested areas are located.

Before any of the above analysis tools can be applied, it is vital to consider the data sets and scale that may be relevant to a particular analysis, along with a few key attributes about each data set, such as source, scale and a currentness reference. There are a myriad of environmental factors and disease ecology components to consider when selecting data sets for an epidemiological analysis. The following is a list of environmental base data sets that were critical to most of the analyses reviewed to determine distribution of a disease: census, soils, vegetation, land use-land cover, climate, terrain, elevation, climate, geology, roads, plantations, watersheds, villages and sometimes various satellite imagery. Within the United States, Advanced Very High Resolution Radiometer (AVHRR), State Soil Geographic Database (STATSGO) and Topologically Integrated Geographic Encoding and Referencing System (TIGER) appear to be heavily utilized, along with Etak teleatlas on a much smaller scale. While this is a list of relevant data sets found in the literature, it is important as part of a GIS planning exercise to determine the data sets that would be most beneficial to answering the question at hand. For example, the geographical extent of an animal habitat, disease distribution, areas of infection, treatment centers, coordination units, etc. should be considered for the epidemiological aspect of the analysis. It is important to note that some of these data sets are not readily available for all countries, but that they do represent the data sets that need to be acquired in order to conduct a good analysis. One of the current problems with studies of epidemiology is that the raw information is lost in reports as aggregate data sets, which are not helpful for the building of a GIS (Mott et al., 1995).

From the literature, it appears that GIS has a lot of growing room related to its functionality determining the distribution, habitat and environmental
factors pertinent to the control of a particular epidemiological concern. With this in mind, it is important to note that all of the power in a good GIS is bound to the data used for analyses. Therefore, it is imperative that as the field of GIS and epidemiology continues to grow, all potential data collectors need to be aware of the necessity to accurately collect data and to check the data before analyses are conducted. Once incorrect data is input, it is hard to catch and update accordingly. The key is that the usefulness of a GIS is directly linked to the reliability of the data collected. Therefore assuming that data collection is done correctly, there is an opportunity for GIS to assist in the implementation, monitoring and maintenance of control programs directed at a specific disease or area. As GIS increases in its usage related to epidemiologic research, it is important to bear in mind the appropriateness of combining data sets, since this will have an impact on any outcomes from the analysis.

In the future, as the interest in GIS is developed, it is recommended that data collection guidelines for the epidemiological community be generated, particularly for endemic countries, where diseases are restricted or peculiar to a locality or region (Merriam-Webster), so that the quality of data collection will be addressed. The focus of most analyses appear to be on the present and the future, which is not bad, but it is also important to not forget about the past by mapping historical environmental and ecological data, when possible. In the end, all of the great GIS analysis and modeling applications are irrelevant without accurate data to support them, so the current focus for research, especially in endemic countries, should be simply compiling a database of relevant information and double checking it. Currently, one of the biggest challenges of GIS analyses is the balance between accuracy of coverage and speed of repeatability and reliability of analysis precision for decision making related to disease control (Hendrickx, et al., 2001). With this said, one of the key concerns for GIS in the future related to epidemiological studies will be where to find the funding to create the necessary data sets, buy the software to get it done and conduct analyses. If funding is available, GIS will be used to support important planning and policy development related to management of specific epidemiological concerns, as it is an intuitive method to display the interaction between several key elements at one time. It is also possible that GIS will assist in maximizing the use of staff and budgets, particularly in developing countries (Mott et al., 1995).

From this review, hopefully it is obvious that GIS can assist not only in mapping disease attributes, environmental factors and other pertinent data sets, but that the greatest impact of GIS will come from future analyses that are directly related to the control of a particular disease by detecting spatial patterns over time. Mott echoes this statement, when he conveys his belief that there is a potential for a complete revision of our current understanding of tropical disease distribution and prevalence, if GIS is extensively implemented to assist in planning and monitoring the control of a tropical disease (Mott et al., 1995). While this statement has a powerful effect, please note that none of this will be possible, unless the collaboration among epidemiologists,
environmental scientists, biostatisticians, and local communities, etc. are continuously reinforced and nurtured. It takes all of these to realize the possibilities of GIS in epidemiologic research, which will assist in implementing, monitoring and hopefully controlling the problem from a landscape perspective.

Annotated Bibliography:


Barnes et al. discuss the role of GIS and landscape ecology to the management of plant viruses in agroecosystems. Problems of viruses in Del Fuerte Valley of Sinola, Mexico are examined. GIS contributed the most to management within this valley by increasing cooperation and information exchange among the key stakeholders in local and regional groups. The result of this new exchange was not only decreased occurrences of disease, but also a great decrease in the use of insecticides. In addition to this, other practices that supported the management principles of avoidance and eradication were set in place. In the end, it is important to note the applications of GIS for today, but also consider the importance of cataloging a historical perspective to maintain the long-term perspective when making decisions.


De Jonge et al. have conducted an interesting analysis on historic eelgrass beds for the purpose of restoration. In this article, maps from 1869 and 1930 are compared with nautical charts to determine the amount of eelgrass that was in the subtidal area at that time. For the purpose of restoration, the answer is simply that there were subtidal eelgrass beds in 1869. This is important to document, since it links the disappearance of the eelgrass beds in the 1930s with wasting disease and confirms that eelgrass existed. The main reasons for a difference in eelgrass percentages between 1869 and 1930 are hypothesized to be deterioration in light conditions due to the closure of Zuiderzee and a fishing technique started in the 1920s involving propeller driven vessels with a higher trawling capacity, which are important to note if restoration efforts are going to be successful in the future.

In this paper, Guerra et al. discuss habitat suitability for the tick species, *Ixodes scapularis*, in the North Central United States. Suitability analysis is done using bedrock geology, soils, topography and vegetation patterns. Each variable was categorized into a positive and negative outcome for tick abundance. The Wisconsin and Illinois area was turned into a grid and the Summarize Zone procedure of ArcView was used to summarize attributes for features in each grid cell. A logistic regression equation was used to generate the probability of the presence of *Ixodes scapularis*. An important aspect of the suitability map created was that vegetation, soil and topography are related and an extreme event in any of these will effect tick survival. The key of this study was to determine the factors that limit tick survival, so that control measures can be assessed based on the risk map created.


In this paper, Hendrickx et al. discuss how GIS can be used as problem solving tools that allow users to analyze data in a multidisciplinary context. For example, a GIS used to study African animal trypanosomosis (AAT) would be useful to examine spatial patterns of the tsetse fly, assess trypanosomosis risk, map clinical disease, livestock biomass, breed distribution, farming systems, and land use. GIS has been used to deal with all of the above issues over the course of the last ten years. However, during the review of current GIS developments and analyses related to AAT, it was determined that GIS has come a long way in the last ten years, but there is still plenty of room for development, particularly related to vector abundance and assessing predictor variables. In the future, the hope is that in time, GIS will assist in forecasting and assessing changes over time of all variables related to AAT. The biggest challenge with the current GIS data and analyses is maintaining a balance by examining a cost: benefit ratio of data that considers the accuracy and coverage of data sets and the speed and repeatability of an analysis versus precision.


In this paper, Kistemann et al discuss the evolution of GIS as an invaluable tool for processing, analyzing and visualizing spatial data. They determine that the best aspect of a GIS for map creation is that the data sets are dynamic and can
easily be annotated to meet the needs of the user. The analysis portion of GIS is viewed as primarily interpolating and modeling of environmental factors relevant to the distribution of diseases within a certain environment. However, they also feel that GIS has a long way to go before it can adequately detect the spatial patterns of disease, primarily regarding the dimension of time. In the end, it is reviewed as a great tool with many possibilities for future development as a tool for epidemiologic research.


In this paper, Kitron et al. discuss the use of GPS, GIS and Remote Sensing as tools for spatial analysis of the epidemiology of vector-borne diseases at the landscape ecology level. GPS, GIS and Remote Sensing are considered new tools for gathering, processing, managing and analyzing data sets related to the epidemiology of vector-borne diseases. This article examines the components of spatial analysis and discusses the importance of scale when looking to make connections between all of the various biological, social and environmental parameters. Several examples of use are discussed in case studies. The key to this article is that it conveys the importance of having a dynamic interface to deal with the dynamic data sets that are critical to demystifying the complexity of developing and applying surveillance and control methods for vector-borne diseases.


In this paper, Mott et al. discuss how diseases are currently managed and how GIS may assist in the future control of tropical disease. Current limitations in data availability and disease management are highlighted. They make a strong case that GIS could fill a very important role in managing tropical diseases more effectively through the consideration of spatial relationships between distribution, factors that affect distribution and health care units in a country. GIS is presented as a way to maximize decreasing health budgets and staff for the management of disease.


In this paper, Vine et al. discuss how GIS technology is applicable to research in epidemiology. They do a great job of discussing and defining all of the different pieces that make up a GIS from data formats, database development and database integration to specialized functions of a GIS. With the basics clearly
defined, they describe three case studies illustrating specialized functions related to epidemiologic studies. In an attempt to clarify and emphasize the function of address geocoding, a case study from North Carolina is highlighted to show some of the hindrances involved with this approach. This paper was particularly interesting, since it had a small section addressing procedures of confidentiality, which is important when conducting any medical study. In their final discussion, the authors review the items to consider when planning to build a GIS and emphasized the importance of collaboration and communication when building a GIS.

Other literature:


Merriam-Webster Dictionary (http://www.m-w.com/cgi-bin/dictionary)