Mapping new techniques in public health: A review of the use of geographical information systems in public health research.

Daren T. Nicholson, M.D., C.M.

INTRODUCTION

For centuries public health officials have used maps in the course of their work. Indeed, the power of cartography was recognized by many of the originators of public health research, including the renowned John Snow. Like these pioneers, modern public health workers rely on maps to better understand the epidemiological variable of "place." When compared to evaluations of the variables "person" and "time," an analysis of "place" may pose more significant challenges to the unsuspecting researcher. Fortunately, one can overcome these obstacles with the use of geographic information systems (software programs, otherwise known as GIS, that allow the user to code and then spatially display various forms of data in a geographic context such as a neighborhood, city, country, etc.). As evidenced by the recent increased use of GIS in public health research, future advances in public health practice will depend on data that is not only current and accurate but also geographically-referenced.

This article is a broad overview of the use of GIS in the public health sector. After covering some foundational information concerning GIS, the article will summarize the types of public health projects that involve GIS technology. Finally, in the discussion section I will contrast the strengths and weaknesses of GIS.

BACKGROUND INFORMATION

History of GIS

Although Foresman traces the pedigree of geographic information systems as far back as the eleventh century, researchers developed the first computerized GIS approximately 40 years ago when computer technology was in its infancy. Generally, GIS historians recognize the Canada Geographic Information System (usually shortened to CGIS) as the first fully-developed GIS. This project stemmed from a land-use policy of the Canadian federal government that emerged during the late 1950s and early 1960s as a means of managing the country's vast natural resources. Shortly thereafter, in the mid-1960s, research into GIS sprung up elsewhere, most notably at Howard Fisher's Laboratory for Computer Graphics at Harvard University. From these initial cells of research, GIS technology has advanced in stride with new innovations in computer and information science.

GIS Components

Every GIS project depends on the interaction of four factors: hardware, software, data and users. A discussion of issues related to the latter two components will follow. Here we will consider the software and hardware requirements of GIS. Given the large amounts of data involved in most GIS projects, and given the complexity of the mapping process, GIS projects require adequate investment in both computer hardware and software. The hardware components needed to complete a typical GIS project include a large hard disk (standard on most PC's currently on sale), adequate RAM and processor speed, a monitor with a minimum pixel resolution of 1024 x 768, a color printer, and possibly a scanner. In addition to these hardware components, project managers should consider installing a high-speed internet connection since the most efficient acquisition of GIS data is often by means of downloading large files from the internet. In terms of acquiring GIS software, public health officials may choose from several off-the-shelf packages (see Table 1). Each product varies by price, functionality, ease-of-use (including documentation), and the amount of data pre-packaged with the software. Thus, organizations have a significant degree of flexibility when choosing the most appropriate GIS software package.

Steps in a GIS Project

Once the appropriate hardware and software have been chosen, most public health GIS projects proceed through the same sequential phases. The initial stages of gathering and organizing the data tend to be the most costly portion of a GIS project. First, one must obtain the relevant attribute data (data that are non-spatial, which therefore could be any type of data deemed relevant to public health). To map attribute data, each record in the attribute database must contain a spatial reference (such as a street address or a set of longitudinal and latitudinal coordinates). Next, the attribute data must be geocoded. The geocoding process links the spatially-tagged attribute data to the corresponding geographic data (such as street segments) that are electronically represented in a geospatial database.. For example, a particular attribute datum representing a case of syphilis would be spatially referenced by the address of the afflicted individual. To geocode the selected datum, the GIS software verifies that the individual's address exists in the geospatial database. If the address in fact exists, then the GIS software will effectively "plot" this datum point on the digitized geospatial map.

When geocoding, one must ensure that all data (in both the attribute database and the geospatial database) share the same scale and projection. The scale correlates the size of an entity (such as a distance) in a spatial database with its actual size. Projection is the process by which the curved, three-dimensional shape of the earth's surface is mathematically transformed so that it may be presented on a two-dimensional map. Mismatched scales and/or projections will lead to significant data misrepresentation. The process of geocoding can be laborious, but, "[a]fter correcting obvious errors [such as abbreviations and misspellings], it is typical to achieve a match rate of more than 90 percent."

Once geocoding is complete, GIS projects shift into the map design phase. Early in the design process, one must create various, overlapping map layers. One layer of a GIS map may display roads, while another layer may render polygons (e.g., territorial or zip code boundaries, etc.), while a third may contain only points (e.g., individual homes, schools, hospitals, etc.). The user creates and manipulates these layers based on her preferences. GIS users may also choose from a variety of themes for their maps. One such thematic map is called a chloropleth map, which displays information about portions of the map using graduated colored infilling. These gradations of color represent variations in the attribute data associated with the predefined regions of the map. Another design approach would be the dot density theme, which displays an individual dot on the map for each bit of raw data. A third possibility is to create a graduated symbol map. Variations in the size of a graduated symbol provide information on the magnitude of the underlying attribute data. For example, a small school symbol on a GIS map may represent a local kindergarten, while a large school symbol would represent a central high school. GIS products may also allow multi-theme maps (for example, a dot density map may overlay a chloropleth map). The user must decide which theme best communicates information about the underlying data.

GIS PROJECTS IN PUBLIC HEALTH

Despite the long history of GIS use in other disciplines, public health has only recently begun to harness the power of this technology. In fact, the earliest Medline articles referenced by the MESH term "Geographic Information Systems" date from mid-2001 (this MESH term was introduced in 2003; thus, some articles prior to 2003 have been re-indexed using this new MESH term, while earlier articles related to GIS continue to be indexed under other MESH terms). In a Medline search conducted by Kaiser et. al in January 2003, the authors acknowledged that "the first public health article using GIS [was] published in 1986, coinciding with the time GIS became available on personal computers; 94 per cent of the 255 articles in [their] search were published in 1995 and later." Since January 2003, Medline has indexed approximately 220 new articles under the MESH term "Geographic Information Systems." Thus, there is a tremendous rate of growth in this area of public health research.

The most common use of GIS in public health is to investigate the geographic nature of disease. As Cromley notes:

> The medical geographic approach to the study of disease has not been the dominant perspective in the United States. Nevertheless, efforts to map the spatial distribution of human cases of disease and the geography of environmental risk…have been made again and again throughout our history.

Multiple articles have been published that review the use of GIS in the study of human disease. Studies usually focus on one of three types of pathological causes: biological (infectious), environmental, or injury-induced. Increasingly, the distinction between "biological" and "environmental" causes of disease has become blurred. Indeed, it is quite difficult to determine how biological agents, environmental conditions, and human behavior interact in pathological processes. GIS may help public health researchers elucidate the complexities of these interactions.

Although most of the articles published on GIS use in public health focus on disease prevention, one can apply GIS techniques to other aspects of public health. For instance, GIS may be used in the management and prevention of humanitarian disasters. Kaiser and colleagues cited recent international conflicts, such as those in Kosovo and Afghanistan, in which emergency relief workers used GIS as a standard tool for planning resource allocation and other logistical concerns. These authors predict that GIS technology will be used to construct early warning systems for emergencies such as famine and drought. Coupled with global positioning systems technology, GIS could prove to be a timely way of analyzing dynamic situations such as a mass migration of refugees.

In their article on public health policy and GIS, Roper and Mays forecast expanded use of GIS to assess and improve the quality and accessibility of health services. For instance, public health officials may choose to map the primary care clinics in their region to analyze the allocation of primary care services. The goal of such a project would be to identify areas of overlapping coverage as well as regions lacking sufficient primary care resources. GIS maps can also aid in the strategic planning process (e.g., choosing a location for a new immunization clinic). Thus, good use of GIS technology can have a significant impact on the delivery of health services.

Although GIS is a relatively new tool in public health, the research projects cited above are evidence that GIS is becoming firmly established as an effective method of doing public health research. In fact, GIS technology will likely change how public health is practiced, including how researchers think about gathering, analyzing and displaying public health data. For instance, GIS may lead to the incorporation of non-traditional public health data (such as the use of juvenile arrest rates in Oregon) into future research projects.

DISCUSSION

The chief strength of GIS is that, when properly designed, maps can be a valuable means of communicating public health information. Like charts and graphs, maps present data in an organized, comprehensible and visually appealing manner. Well-designed maps can transmit the complexities of epidemiological information to nonepidemiologists such as policy makers, community members, or even other health care workers. Since successful public health initiatives often depend on these "laypersons," use of effective communication techniques such as GIS maps are of paramount importance to public health officials.

In spite of the aforementioned strength of GIS as a public health tool, the use of this technology has a few drawbacks, as numerated in Melnick's textbook on GIS in public health. One of the strengths of GIS, namely their ability to provide detailed geographic information, leads directly to a limitation of GIS usage: if not adequately monitored, GIS maps may disclose confidential information. A published address can reveal the identity of an individual, even if the individual's name is not divulged. Fortunately, these types of disclosure are rare. More frequently, individuals may be identified indirectly in situations when the population and/or geographic area under investigation are small. In such circumstances, public health officials should choose to disseminate only aggregated data (e.g., summary statistics on a county level) to prevent disclosure of sensitive information about individuals. That being said, even aggregated data may raise questions about confidentiality since disclosure of information on groups, such as a neighborhood, can have a negative financial impact on that community (e.g., decreased property taxes, increased insurance premiums, etc.). Considering issues of confidentiality before beginning a GIS project may prevent wasted hours creating maps that cannot be published due to breaches of confidentiality.

Poor data availability and reliability may also confine a GIS project. One of the pitfalls of many GIS projects is a poor match rate between addresses in the geospatial database and those in the attribute database. Poor quality data in either database may lead to such a mismatch. By excluding non-matched attributional data, the results of subsequent analyses may be subject to a selection bias if the remaining (selected) data do not fully represent the entire set of attribute data. Even suitably matched attribute data may have other qualitative constraints. Thus, meta-data (e.g., the source, quality and currency of the data) are indispensable elements in any GIS project.

Perhaps the most limiting factor in a given GIS project is the human component. Users may easily create maps of public health data if all the necessary tools are at hand (good data and an adequate GIS); however, proper analysis and interpretation of the data require a trained public health worker. In fact, Melnick avers that "[m]ost geographic analyses assessing whether there is an association between geography and health outcome will find one. Usually, however, outcomes such as cholera will cluster geographically because of underlying population characteristics, not because of the geography itself." For instance, when mapping crude breast cancer rates, clustered data would usually reflect the geographic distribution of sex and age (breast cancer being more common among women, especially older women) not the geographic distribution of breast cancer risk. Mapping age-adjusted breast cancer rates would provide better information on the geographic patterns of breast cancer in a given community. Without considering these confounding factors, users are likely to project personal experiences and biases onto the map and thus unwittingly infer that the health data are clustering because of their own hypotheses about its geographic distribution. Thus, it is essential that well-trained public health workers are involved in the analysis and interpretation of GIS output.

Ironically, the robust capabilities of GIS may lead to poorly designed maps if users are not careful. The seemingly limitless design possibilities of current GIS software can be overwhelming. Users must pay close attention to their design choices or risk unintended results. For instance, to the reader of a chloropleth map, the size of a uniformly colored area may appear to be as important as the difference in colors among various regions. Thus, large, sparsely-populated areas might erroneously seem to dominate smaller, more densely-populated regions. To easily navigate the myriad design possibilities of GIS software and thus overcome this shortcoming of the technology, users must obtain adequate training.

CONCLUSION

The current flurry of GIS related public health research may prove to be a short-lived fad; but as technology continues to progress, it seems probable that public health researchers will continue to find uses for GIS in their work. Indeed, the continued advance of internet technology should fuel the expansion of GIS related public health research. Further integration between GIS and the Web will continue to eliminate many of the data-silos that were pervasive before the dawn of the internet-age. Given the motivating force of technology innovation coupled with an emerging emphasis in public health on the spatial nature of data, GIS use in public health should continue to expand over the foreseeable future.

Table 1 – GIS Software Products*

Producer	Software Product
ESRI	ArcView GIS
Autodesk	Autodesk World
Intergraph	Geomedia
MapInfo	MapInfo
Caliper Corp.	Maptitude
CDC	EpiMap
Caliper Corp. & Dept. of Housing and Urban Development	Community 2020

*Table adapted from