A Review and Framework for Categorizing Current Research and Development in Health Related Geographical Information Systems (GIS) Studies

A. K. Lyseen1, C. Nøhr1, E. M. Sørensen1, O. Gudes2, E. M. Geraghty3, N. T. Shaw4, C. Bivona-Tellez5 on behalf of the IMIA Health GIS Working Group

1 Department of Development and Planning, Aalborg University, Aalborg, Denmark
2 Department of Spatial Sciences, Curtin University, Australia.
3 University of California Davis, Division of General Medicine, California, USA
4 Azusa University, Sault Ste. Marie, Ontario, Canada
5 Azusa Pacific University, California, USA

Summary

Objectives: The application of GIS in health science has increased over the last decade and new innovative application areas have emerged. This study reviews the literature and builds a framework to provide a conceptual overview of the domain, and to promote strategic planning for further research of GIS in health.

Method: The framework is based on literature from the library databases Scopus and Web of Science. The articles were identified based on keywords and initially selected for further study based on titles and abstracts. A grounded theory-inspired method was applied to categorize the selected articles in main focus areas. Subsequent frequency analysis was performed on the identified articles in areas of infectious and non-infectious diseases and continent of origin.

Results: A total of 865 articles were included. Four conceptual domains within GIS in health sciences comprise the framework: spatial analysis of disease, spatial analysis of health service planning, public health, health technologies and tools. Frequency analysis by disease status and location show that malaria and schistosomiasis are the most commonly analyzed infectious diseases where cancer and asthma are the most frequently analyzed non-infectious diseases. Across categories, articles from North America and Europe dominate, and in the category of spatial analysis of diseases an equal number of studies were performed in Asia.

Conclusion: Spatial analysis of diseases and health service planning are well-established research areas. The development of future technologies and new application areas for GIS and data-gathering technologies such as GPS, smartphones, remote sensing etc. will be nudging the research in GIS and health.

Keys

Epidemiology, geographic information system, public health, cluster analysis, spatial analysis

Introduction

Place is the central element in all spatial analyses and varies in both size and shape depending on the context. For more than two thousand years, since Hippocrates’ On Airs, Waters, and Places, place has played an important role in understanding health and disease [1]. Historically, maps have been the primary source for storing and communicating spatial information. Maps have been used to describe geographic health and disease phenomena, which Filippo Arrieta was the first to do in 1798 to describe cases of Yellow Fever in New York [2-3]. The most famous example of disease mapping is John Snow’s mapping of cholera cases in 1854 in London [4]. Maps have traditionally been presented on paper, either individually or in atlases. The development of computers has extended the concept of maps significantly to embrace maps displayed electronically as static maps, animated sequences of maps or interactive maps.

In the last two decades, the possibilities for mapping and for spatial analysis of disease patterns have changed dramatically, as computer power has increased and Geographic Information Systems (GIS) have emerged as individually accessible software, allowing for more widespread, complex and comprehensive analyses than previously. Such advances have made it possible for medical geographers, and others, to seek answers to questions that were previously overly complex and unfeasible. Through GIS analysis it is possible to understand why things are located where they are and, in combination with health and other sciences, how they are related. Obtaining disease and health data has been made easier by low-cost global positioning system (GPS) units and the improvement of the quality of Remote Sensing (RS).

In the course of the last centuries innovators such as Arietta, Finke, Seaman [2] and others [5-10] have mapped disease and measured incidence and prevalence rates, but what about measuring health? Measuring health or well-being includes cultural, social, environmental, biological and physical indicators that are often related to a geographical aspect, which Albright and Ramasubramanian [11] utilized in developing their Geographic Index of Relative Wellbeing. However, these indices for measuring health quality are based on society’s individual cultural, social, environmental and biological characteristics and are therefore seldom applicable for uses other than the narrow purpose for which they were developed [12].

GIS can be used to manage the modelling and mapping of disease; to develop new hypotheses in a geographic context; to analyse and predict future disease risks; and to undertake location/ allocation analysis of the distribution of services and resources. Additionally, the advances and increased affordability in DNA sequencing has added another layer of complexity and potential insight [13].
GIS has, therefore, become an essential tool for health professionals who work with health data that contain a spatial reference. GIS has become increasingly available in the last two decades, and the implementation of GIS in health research has developed along the way. Figure 1 shows the number of scientific articles on GIS in health published each year from 1991 to 2011. In 1991 only eight articles were published, in 2000 there were 78, and in 2011 more than 450 articles were published and indexed in either Scopus or Web of Science with the keywords ‘GIS’ and ‘health.’ GIS was added as a MeSH term in 2003 demonstrating its growing importance in the health literature.

Data accuracy, correctness and completeness are key elements that affect our ability to effectively use GIS to undertake analysis of health issues. Collecting geographical health data has previously been limited, but the development and enhancement of technologies such as Global Positioning Systems and remote sensing technologies have made it easier to obtain geographical health data. In fact, in the USA, on May the 1st 2000 Bill Clinton removed the selected availability of GPS signals, which improved the accuracy of the spatial references that civilians could receive tremendously [14]. Despite the significant improvement in technologies to obtain geographic data, we still need to be able to geocode patient data to be able to undertake effective geographical analyses. This imposes a major challenge to health information systems in collecting data with adequate structure and granularity, ensuring reliability and validity of the relevant health data, and still maintaining appropriate privacy and security for patient records. Further, even when such data is collected it is often held “hostage” under the guise of privacy legislation requirements preventing ready access to key data sets.

Despite the challenges, research comprising health geography and GIS has expanded worldwide. The objective of this study is to review the literature and build a conceptual framework for the published research on health and GIS. The first step in building this framework is the identification and analysis of available literature. New and established research areas will be identified and classified for inclusion so that the framework can be used in strategic discussions and prioritizations in research environments working on GIS applications in health science. In this study the focus is on the applications of GIS and not on the technical and standardization aspect of GIS systems. Furthermore we focus on the public health area and not on environmental health.

**Methods**

We acknowledge that there are several standards and guidelines for writing reviews (Institute of Medicine, Cochrane, PRISMA etc.) and this paper selected the most relevant components among them. The identification of relevant literature was limited to journal articles, conference papers and reviews in the library databases Scopus, which includes the entire Medline library, and Web of Science published in the years 1991 to 2011. Only articles published in English were included. The search was limited based on the keywords designated to each article.

Papers covering veterinary health, crop or other areas of plant health, building or structuring health, and the health of the environment and ecosystems were identified in the original search, but excluded from further study, as the focus of this framework is on human health and disease as they relate to geography and GIS.

The papers included in this framework were selected and analyzed in a four step process:

**The 1st step:** The library database Scopus was used to determine the volume and characteristics of the literature and define the keywords strategy and search terms. The search terms applied are shown in table 1.

**The 2nd step:** The library database Web of Science was searched with the same search terms as in table 1. Articles with topics that matched the search terms and were not already included in step 1 were added to the list of references.

**The 3rd step:** The reference lists of each of the papers selected during steps 1 and 2 were examined. Articles whose title coincided with any of the search terms were deemed relevant and also included in the study.

**The 4th step:** All the abstracts of the identified papers were read. Using a grounded theory-inspired approach, and passing through several iterations, they were grouped into categories based on their main focus. The key papers, that were felt to speak to fundamental issues, and the papers that were difficult to place in a category were read in their entirety. The categories were treated individually, but in reality papers constituting the categories are often related and it was consequently difficult to

![Fig. 1 Articles published each year under the search term ‘GIS AND health’.](image-url)
achieve complete exclusiveness among the categories and, hence, some articles occurred in more than one category.

It should be noted that remote sensing technologies were added to the search because they constitute an important issue when discussing GIS. GIS needs accurate data to analyse in order to give added value, and are included here to show their significance in the categorization.

Results

The first step in the literature search in the library database, Scopus, yielded 562 references published in the years 2000 to 2011. The search in Web of Science added 144 references, and searching the list of references in these 706 papers added another 159 references, creating a total number of 865 papers.

The first iterations of reviewing the articles identified a number of key areas as described below:
1. The basic principles of disease geography focus on describing, exploring and analyzing the incidence and prevalence of disease in a spatial context – an epidemiologic surveillance approach, and
2. Disease surveillance which can be divided into two coherent elements: disease mapping and disease modelling [15].

Disease mapping describes the past and present distribution of disease; past and present spread and transmission of disease; disease incidence and prevalence rates for monitoring health quality; surveillance of environmental hazards; historical trends in disease; combining epidemiological, environmental and geographical information to generate new hypotheses; and evaluating the outcome of interventions and disease preventive activities. Disease modelling takes disease mapping a step further by using past and present disease information to predict future disease outbreaks and transmission patterns; analyze the risk of disease; predict environmental factors that promote disease; test hypotheses; determine disease hotspots for intervention or prevention; and monitor the drivers of epidemics.

The geography of disease is intertwined with the geography of the health care system [16]. The close link between the geography of disease and that of the health care system is significant and has implications regarding the allocation of resources where they are most needed [17]. The geography of the health care system is resourcing the supply of care where it is needed; planning the service structure; allocating staff and resources; allocating and dispatching emergency services; minimizing disparities and inequities in the access to services and facilities; and analyzing the effect of distance to services on the utilization of said services and on health outcomes.

The geography of disease and the health care system cover the surveillance of disease and illness and the planning, provision and management of services to cope with disease and illness – the curative or management aspects of disease and health. Apart from the curative aspects, public health aspects such as health promotion and preventive activities are part of the research carried out within health geography and GIS. Health promotion and prevention are related to reducing the chemical, physical, infectious, behavioral and mental impacts that humans deal with every day and to promoting a healthier lifestyle via, for instance, physical activity and better nutrition.

During the review iterations four distinct categories crystallised: spatial analysis of disease, spatial analysis of health service planning, public health, health technologies and tools. The four categories are defined as follows:

- **Spatial analysis of disease** embraces disease mapping and modelling; geographical epidemiology such as disease detection, prediction, surveillance and monitoring; environmental epidemiology such as causality and risk analysis, disease transmission, and the analysis of disease patterns.
- **Spatial analysis of health service planning** focuses on spatial analysis for planning, management, delivery, provision, accessibility and utilization of health care and emergency facilities.
- **Public health** comprises spatial analysis for promotion, preventive and rehabilitation activities as well as spatial analysis of health outcomes.
- **Health technologies and tools** concentrates on technologies for collecting health data such as GPS, remote sensing, and personal peripheral devices like accelerometers and heart rate monitors as well as health data manipulation tools like geocoding and data mining.

### Table 1

<table>
<thead>
<tr>
<th>Set</th>
<th>Search terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Health* OR Disease* OR Illness* OR Medicine OR Medical*</td>
</tr>
<tr>
<td>1</td>
<td>[A] AND [B] AND Epidemiology</td>
</tr>
<tr>
<td>3</td>
<td>[A] AND [B] AND Disease Cluster OR Disease Pattern</td>
</tr>
<tr>
<td>4</td>
<td>[A] AND Disease Mapping OR Disease Modelling</td>
</tr>
<tr>
<td>6</td>
<td>[A] AND [B] AND Planning OR Management OR Delivery OR Provision OR Accessibility OR Utilisation</td>
</tr>
<tr>
<td>7</td>
<td>[A] AND [B] AND Location Analysis OR Allocation Analysis</td>
</tr>
<tr>
<td>8</td>
<td>[A] AND [B] AND Health Preventive OR Health Prevention OR Prophylactic OR Health Promotion</td>
</tr>
<tr>
<td>9</td>
<td>[A] AND [B] AND Rehabilitation OR Health Outcome</td>
</tr>
<tr>
<td>10</td>
<td>[A] AND [B] AND Telehealth OR Assisted Living</td>
</tr>
<tr>
<td>11</td>
<td>[A] AND [B] AND Geocoding OR Data Mining</td>
</tr>
</tbody>
</table>
Table 2 shows the categories, the number of articles in each category and the primary keywords describing the content.

The publications found were analyzed by publication year to show the chronological development from 2000 to 2012 in the four groups identified – see Figure 2. The publications in the area of spatial analysis of disease dominated the entire period, and during the last year the number of publications in this area almost doubled from 62 to 100. The other three categories have seen up to 20 publications per year. During the last year of the study period the category of spatial analysis of health services planning experienced the same relative increase as spatial analysis of disease while health technology and tools increased by more than 100% (12 to 32 publications per year).

The articles within the category ‘spatial analysis of disease’ were further analyzed for relation to infectious diseases and non-infectious diseases (see Table 3). The articles in these two areas contained 367 references, of which eight articles focused on mortality in general.

Among the articles on infectious diseases malaria was the most common disease undergoing GIS analysis with 40 articles being published on this alone. Thirty-nine articles dealt with Schistosomiasis, a parasitic disease caused by a worm contracted through improper water and sanitation facilities. Dengue fever, HIV and Leishmaniasis occurred 25, 22 and 15 times respectively (see Table 4).

The most common non-infectious disease among the GIS and health articles was Cancer with 31 references. Asthma had about half the number of articles and cardiovascular diseases, diabetes and obesity had only eight, seven and five articles focusing on these diseases (see Table 5).

Given the geographic nature of our work it was appropriate to examine what continent the different articles related to. Table 6 shows the geographic distribution of the 865 studies. Since eight articles concerned more than one continent, the total number of studies represented 873. In the categories spatial analysis of health services planning and health technology tools there are many studies that do not relate to a specific continent because of a generic approach and content so these were categorised as ‘No specific continent mentioned’. The relative proportion of published studies by continent and category is indicated as a percentage in brackets.

### Table 2
The number of articles and the primary keywords for the four categories into which the results have been divided.

<table>
<thead>
<tr>
<th>Number of articles</th>
<th>Spatial analysis of disease</th>
<th>Spatial analysis of health service planning</th>
<th>Public health</th>
<th>Health technologies and tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epidemiology, surveillance, prediction, detection, monitoring, disease cluster, disease pattern, disease mapping, disease modeling, disease transmission</td>
<td>503</td>
<td>307</td>
<td>126</td>
<td>176</td>
</tr>
</tbody>
</table>

### Table 3
Number of articles in the category “Spatial analysis of disease in relation to infectious and non-infectious disease.”

<table>
<thead>
<tr>
<th>Type of Disease</th>
<th>Number of Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infectious</td>
<td>248</td>
</tr>
<tr>
<td>Non-infectious</td>
<td>119</td>
</tr>
</tbody>
</table>

### Table 4
The five infectious diseases most commonly analyzed

<table>
<thead>
<tr>
<th>Disease</th>
<th>Number of Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malaria</td>
<td>40</td>
</tr>
<tr>
<td>Schistosomiasis</td>
<td>39</td>
</tr>
<tr>
<td>Dengue fever</td>
<td>25</td>
</tr>
<tr>
<td>HIV</td>
<td>22</td>
</tr>
<tr>
<td>Leishmaniasis</td>
<td>15</td>
</tr>
</tbody>
</table>

### Table 5
The five non-infectious diseases most commonly analyzed

<table>
<thead>
<tr>
<th>Disease</th>
<th>Number of Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer</td>
<td>31</td>
</tr>
<tr>
<td>Asthma</td>
<td>14</td>
</tr>
<tr>
<td>Cardiovascular diseases</td>
<td>8</td>
</tr>
<tr>
<td>Diabetes</td>
<td>7</td>
</tr>
<tr>
<td>Obesity</td>
<td>5</td>
</tr>
</tbody>
</table>

Fig. 2 Articles published each year within the four main categories.
The following sections describe each category in more detail. Only the 196 most significant references are mentioned, but the result of the entire search is available here: http://www.schattauer.de/fileadmin/assets/zeitschriften/IMIA_Yearbook/A_complete_list_of_literature_found_for_the_article.pdf.

Table 6 The number of articles in the four categories in relation to what continent they are concerned.

<table>
<thead>
<tr>
<th>Continent</th>
<th>Spatial analysis of disease</th>
<th>Spatial analysis of health service planning</th>
<th>Public health</th>
<th>Health technologies and tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>82 (16.0%)</td>
<td>35 (11.4%)</td>
<td>4 (3.2%)</td>
<td>10 (5.4%)</td>
</tr>
<tr>
<td>Asia</td>
<td>139 (27.2%)</td>
<td>33 (10.7%)</td>
<td>6 (4.8%)</td>
<td>22 (11.8%)</td>
</tr>
<tr>
<td>Europe</td>
<td>64 (12.5%)</td>
<td>36 (11.7%)</td>
<td>20 (15.9%)</td>
<td>27 (14.5%)</td>
</tr>
<tr>
<td>N. America</td>
<td>132 (25.8%)</td>
<td>98 (31.9%)</td>
<td>44 (34.9%)</td>
<td>41 (22.0%)</td>
</tr>
<tr>
<td>Oceania</td>
<td>14 (2.7%)</td>
<td>19 (6.2%)</td>
<td>8 (6.3%)</td>
<td>2 (1.1%)</td>
</tr>
<tr>
<td>S. America</td>
<td>60 (11.7%)</td>
<td>4 (1.3%)</td>
<td>2 (1.6%)</td>
<td>9 (4.8%)</td>
</tr>
<tr>
<td>No specific continent mentioned</td>
<td>20 (3.9%)</td>
<td>82 (26.7%)</td>
<td>42 (33.3%)</td>
<td>75 (40.3%)</td>
</tr>
</tbody>
</table>

Spatial Analysis of Disease

The geography of disease is the most basic theme in GIS analyses in health, as knowledge of spatial disease occurrence, incidence and prevalence are fundamental for planning health care services, for health promotion activities and health prevention activities.

As identified earlier, the surveillance of disease consists of two elements: disease mapping and disease modelling: Disease mapping can include illustrations of the distribution of a disease-infected population in the form of incidence or prevalence rates for a given area or more complex models that mimic real world environments by combining multiple data sources that might be considered non-traditional in the health field, like the presence of highways or green space.

Mapping of disease data can lead to discoveries of patterns; ranging from clusters, to uniformly spaced, to randomly distributed data, and including concepts such as how data are dispersed around a center (compactness) or how they trend in a particular direction. Analysis of disease distribution and the recognition of patterns are scale dependent and individual for each analysis [18]. Disease distribution can be mapped on a global [19-20], national [21] or local scale, i.e. a province [22] or a city [23].

Mapping disease patterns in correlation with other spatial information can be instrumental to examine the cause of the pattern by identifying contributing factors and thereby strategically targeting intervention efforts to reduce or remove any disease-promoting factors in the places where they are most prominent [4], instead of just targeting the whole population.

Analyses of disease distribution and patterns are used frequently within health science research; for example, spatial distribution analysis of incidence and prevalence rates for different ethnic population groups [21], distribution of disease mortality [24], mapping and analysing hotspots for disease occurrence [25-28], maps of transmission hotspots based on geographic information of highly crowded areas [29] or analysis of influenza incidence clustering [30].

Analyses of spatial clusters are based on Tobler’s first law of geography stating [31] that things which are closer to each other are more alike than things that are further apart. Tobler’s law has been used as the basis for analysis of disease, as people who live next to each other tend to be exposed to the same factors that promote certain diseases; hence, the analysis of disease patterns and the hazards of disease clusters can help determine the causality for disease outbreaks. Diseases are caused by exposure to different hazards. Jarup [32] showed this well in his analysis of the spatial
distribution of prostate cancer in relation to environmental carcinogens. He mapped the exposure to chemicals in relation to disease occurrence revealing spatial patterns and relationships [19]. Another example is Fischer’s [33] analysis of the spatial distribution of leprosy in relation to the spatial distribution of risk factors. In addition, Wilhelm [34], among others [35-40], analyzed the relation between air pollution and disease, and Chellini [41] analyzed the hazards of living near a sewage plant in relation to the occurrence of lung and laryngeal cancer incidences.

Analysis of spatial hazards such as air pollution can be a challenging task that may include setting up a fine net of pollution monitoring stations. In a GIS, the task may be simplified by focusing the analysis of data on information from a few measuring stations combined with knowledge of factors such as terrain and wind conditions that can provide a smooth, continuous estimate of exposure values between the monitoring stations.

Maps of disease information over time can provide insight into how a given disease behaves and help predict the development of this disease and future needs for intervention and medical facilities. Temporal disease mapping can identify trends over time [42–44] – when the occurrence increases or decreases in specific areas [30, 45]; moves in a certain direction [46, 47] – or is a periodically recurring pattern (cycles).

Temporal analysis aids our understanding of the etiologic drivers of diseases and limit disease-promoting elements, and for evaluating the efficiency of interventions and testing hypotheses for disease reduction.

GIS can be used to test, or generate, hypotheses in either a theoretical environment or actually based on information from the real-world environment, which is what Wagner et al. [48] did. He used GIS to test the hypothesis that elevated groundwater uranium levels and more frequent groundwater use increased cancer incidence. Furthermore, Beyer et al. [49] developed an open source GIS application that encourages community participation and contains information about colorectal cancer in Iowa, thus using the participants’ local geographic knowledge to promote hypothesis generation about colorectal cancer risks and interventions that might reduce the risk.

Modelling of potential future disease occurrence or the early detection of epidemic outbreaks through surveillance of global disease cases is an area under development. Modern disease outbreak monitoring systems such as the Global Public Health Intelligence Network GPHIN [50], Healthmap [51], and BioCaster [52] are, among other things, using online information sources to update their disease maps in real-time. Fleming et al. [53] developed a specific early warning GIS tool to identify suitable environmental conditions for cholera outbreaks in South Africa. Additionally, Lewis et al. [54] analyzed signs, symptoms and diagnostics of disease, gathered on a daily basis and used for monitoring and evaluating disease outbreaks to compare with historical and seasonal patterns in an attempt to detect major disease outbreaks at an early stage. The disease outbreak monitoring system is a great tool for illustrating global patterns of disease, whereas the more specific models for predicting local outbreaks are valuable tools, as these are customised for their region’s unique environmental factors and as such are a better tool for early detection disease epidemic outbreaks.; essential in order to focus health resources and limit the spread of disease. The California Department of Public Health publishes an interactive web map related to Healthcare Associated Infections which exemplifies this use of GIS to limit the spread of disease: It can be seen at http://www.cdph.ca.gov/Programs/Pages/MyHospital411Infections.aspx#/@6/37.546/-119.202?tab=CDI (Last accessed 23rd May 2014)

Modelling geographic areas with a higher risk of disease via temporal and spatial prediction is a valuable tool for limiting future disease occurrence and spread. Multi-criteria decision analysis of causal factors or relationships can identify areas at high risk of disease occurrence by incorporating knowledge of environmental or socioeconomic factors. Examples in the literature mapping the habitats of the host snails for schistosoma japonicum disease [55], mapping elevation height and low population density areas as predictors of a high risk of child mortality [56], analyses of how temperatures affect dengue fever distribution to predict future endemic areas as a result of climate changes [57] or analysis of ecological transformations caused by floods, changes in water resources and climate changes for predicting future infection risks [58].

Relationship analyses constitute another method for identifying high risk areas for disease; for instance, Nakaya et al. [59] developed a model for predicting local HIV epidemics in Japan via geographic variation in population growth and spatial relationships in HIV transmission, and Khormi & Kumar [60] are modelled risk areas of dengue fever on the basis of the relationship between dengue fever cases and socioeconomic factors (such as: population numbers, population density and neighborhood quality).

Spatial Analysis of Health Service Planning

The primary elements in health care systems’ are the patients, clinicians and facilities. They all have a geographical location, and the co-location of all three elements constitutes the highest opportunity for successful management or cure of disease. The absence of planning of services in relationship to the populations served is evident across the globe creating disparities in access for segments of the population. In reallocating health care services often results in a compromise between the population distribution, treatment needs and available resources (equipment and personnel). Incorporating GIS techniques enables for service optimization for population served by incorporating spatial distance, demographic information [55, 56] and capacity [58, 61, 62].

There are multiple methods used for calculating spatial distances. The simplest is the Euclidian (straight line) distance although fast performance and minimal impact on the results in an urban environment but in rural areas can cause errors due to its reliance on aggregating the population in an area to a single point for the entire area population.

A more accurate approach uses a street-level reference datasets that can measure actual distances and travel times which may be especially important for areas where terrain and/or traffic can significantly increase travel times even when distances appear relatively short [57-64].
Apart from location-allocation analysis, parameters for supply and demand are often analyzed, when health care facilities are planned [65-67]. An example of an analysis of the need to plan health care facilities includes Ong et al. [68], who use a database of cardiac arrests to find patterns in the occurrence of incidents and locate defibrillation devices in accordance to the findings. Using GIS to locate optimal places for equipment has also been used in planning the allocation of emergency services; based on geography, population and the projection of needs, to minimise the response time for both ambulances and helicopters [69, 70].

GIS has many application areas within health care planning, but no matter which model or method is used, a compromise between several elements is required. Policy-makers and government officials have to make the decision, and this is where GIS can play an important role as a spatial decision support system, by informing the decision-makers of the pros and cons of different solutions [71-74]. An example of the use of GIS to reflect on the consequences of health care planning by Brabyn [75], who used GIS to calculate the percentage of the population who, as the result of a proposed health reform, must travel for more than one hour to reach an emergency department. Other examples would be the impact of closing a clinic in a region on service resource allocation and access. GIS contains a wide variety of tools for conducting relationship analysis, which helps broaden our insight into the factors that influence the utilization of health care facilities and, perhaps more importantly, health outcomes. Higgs [76] stated, “Research is also needed to examine the potential relationship between distance, utilization patterns and health outcomes.”

Utilization is closely linked to accessibility, which can be measured with a variety of methods. Brabyn et al. [79] and others [80-84] have researched methods for measuring accessibility by comparing different methods. Methods compared by Brabyn et al. [79] include measuring the ratio between patients and medical staff in an area, using drive time and distance analysis to determine the shortest path in time or distance between patients and medical staff and complex network analysis that takes the capacity of the medical staff into account, when calculating the shortest time/distance for patients. The methods are different in terms of accuracy and calculation time and are therefore applicable in different analysis conditions, such as the size of geographic area, the number of patients and medical staff and the complexity of the road network. Careful consideration of the advantages and drawbacks of each method is essential to make an adequate analysis of each area.

Spatial modelling aids in planning to model and predict utilization of services and explore the impact of locating these facilities in different places. Other key factors to consider are how various population groups’ utilization of health care facilities differently. Research into the association between human behavior, geographic accessibility and the utilization of health care has been carried out by Arcury [77] and Chang [78] among others. One factor, in particular, has had an impact on the utilization of health facilities: distance [79, 80]. Higgs [76] examined different spatial analytical approaches to investigate the influence of distance on usage of health services. Other factors’ influence on health service utilization has also been examined. Leung et al. [81] state that factors such as socioeconomic status and regional density have a more recognizable effect on the utilization of health services. Environmental factors are also believed to have an impact on utilization of health services [82], along with physical structures such as rivers and borders [83].

Combining usage of health services with accessibility estimations, theoretically actual accessibility can be derived, as demonstrated by Loh et al. [89], who used patient databases to measure actual accessibility as defined through usage, which was then compared to the potential accessibility calculated on the basis of the shortest path between patient and hospital for example. Others compare the actual distance between patient and health facility with the perceived distance as a measure for the perceived accessibility; hence, the actual accessibility is not merely limited by the actual distance, but also the perceived accessibility, which is individual and affected by factors such as access to public transportation or ownership of a car. Perceived accessibility is concerned with the actual use of the health services in question, as this could lead to better health. Hence, research into associations between accessibility and health outcome have been carried out by several researchers [90-93], focusing on the association between heart disease mortality and geographic access to hospitals in one case [94] and survival from cancer associated with patients’ travel time to health care in another [95].

Inequalities in access to health a service are inevitable, but analyses providing information that lead to furthering our understanding of these inequalities can be used in the planning of health care facilities. Researches into inequalities in terms of accessibility have been conducted by several researchers [96-102].

Public Health

Public health is defined by the Institute of Medicine as “fulfilling society’s interest in assuring conditions in which people can be healthy” [103]. Every day humans are exposed to a fluctuating amount of hazardous impacts on their health in the environment in which they live and travel. The amount and type of hazardous impacts each human encounters differ from area to area, but determinants such as prosperity, educational attainment, technological advancement, health norms, politics, religion, gender, age, occupation, genetics and many more also have an influence.

Health agencies and governments implement health preventing the globe in attempts to reduce the adverse effects of environmental factors. Targeting health prevention activities strategically is crucial, as Alcaraz et al. [104] establish in their study for cancer prevention and control. GIS can provide a deeper understanding of the community and which segments can best be reached with localized interventions, [104]. Miner et al. [106] analyzed the community cancer risk in Alabama, using GIS to target prevention programs. Evaluating preventive activities and health outcomes via GIS has not been widely adopted [102]. Only one article using GIS, Gikandi et al. evaluated
the effects of a malaria strategy in Kenya promoting the use of insecticide-treated nets for pregnant women [107]. They concluded, among other things, that women living further away from the antenatal clinics were less likely to use the nets, even five years after promoting the issue [107].

The environmental exposures can be divided into four groups: chemical, physical, mental or infectious. Chemical exposures can occur through emissions from traffic and industry; drugs, alcohol or smoking; contaminated earth or water; pesticides in drinking water and vegetables; poisoning; allergy-provoking chemicals; and lack of important vitamins and minerals in daily nutrition. Measuring air pollution and its effects on the public health has been researched eagerly in the past decade. The majority of these studies measure air pollution using methods [36, 108] that are based on the residence of the individuals in question [35, 37, 38, 109, 110] and less often on the individuals' exposure outside their residence, i.e. in commuting and at the workplace. Isakov et al. [111] presented a method for combining several models for measuring air pollution to get a better estimate of personal exposure, and Lindgren et al. [37] used a questionnaire to obtain self-reported exposure at a personal level. The effect that air pollution has on human health can be associated with allergy [35, 37], respiratory diseases [35, 109], cardiovascular disease [110] and mortality [38, 110].

Research on lead poisoning in children has, in several cases, indicated an association with older housing [112-114], and GIS has been used to determine areas at risk, leading to preventive actions [115] by screening the children in the affected areas [113, 116, 117], lead hazard remediation of houses [117] and the development of healthier neighborhoods through increased walkability and green spaces [114].

Exposure assessments, as in the lead prevention or air pollution cases, are useful applications of GIS for improving the health of vulnerable groups, and they can be applied to prevent or minimize the adverse effects of chemical exposures that humans encounter. Other examples of exposure assessments via GIS have measured the trichloroethylene levels in water and its association with neuro-behavioral deficits [118] and the association between nitrate content in drinking water and breast cancer [119]. Exposure assessments are also applicable to other types of impacts, such as the physical, mental and infectious.

Exposure assessments of physical hazards; such as different types of radiation (UV or electromagnetic waves) [120-125], weather changes (extreme heat or cold) and their association with increased probability of mortality [112-128], cardiovascular diseases [129], melioidosis cases [130], and impacts on disease transmission [131] have all been studied. Exposures to noise and light have been mapped to measure the effect on humans [132-135] and associations with specific diseases [136-140], e.g. hypertension and insomnia. Mapping hotspots of traffic accidents, involving vehicles, cyclists and pedestrians provides opportunities for city planners to target interventions to the highest risk areas aimed at preventing accidents and physical injury [141-143].

Measuring the exposure of mental impacts through GIS for prevention and promotion has limited research coverage currently and what does exist concentrates on stress, noisy crowds, relationships to family and friends, society and fear of being assaulted physically or verbally. This is due to the inherent difficulties in obtaining mental health data. Zhang et al. [144] investigated the relation between individual and neighborhood education levels and mental health, which could help target preventive activities to areas with low levels of education. Prevention of depressive episodes for individuals with bipolar disorder can be supported when relevant, due to GIS and GPS monitoring of symptoms. For example, when changes in a patient’s daily routine [145] are observed (such as failing to attend clinic for medication administration) interventions can be targeted by Mental Health Support workers to assure that the individual remains stable.

The built environments impact on disease transmission of viruses, parasites and bacteria transmitted from other humans, insects, fish or animals can be related to the quality of buildings; the materials inside the buildings, as these favor different diseases; the quality of water; and whether the place one has chosen to live has a high or low density population. Attempts to prevent the transmission and occurrence of infectious diseases have utilized GIS to map the distribution and incidences of diseases [28, 146-149] and analyze the environmental risk factors promoting the spread of a given disease, such as animals [28, 147, 150] or water [151].

Health promotion is about encouraging people to live a healthy life by increasing physical activity, consuming nutritious foods and increasing happiness through social community cohesion and equality, which positively affect the human resistance to disease. A great amount of research promoting physical activity and thus improve the general health has been done – from analyses of walkability in urban areas, as a function of the environment [152-156], to assessing the suitability for walking and cycling [157] and using GPS and accelerometers [158, 159] to measure individuals’ physical activity levels’ association to the housing environment and its impact on the individuals’ health and wellbeing. Research has shown the following environmental factors influence the physical activity: [160]; proximity to green areas, vegetation and recreational areas [152, 161], streets’ and roads’ suitability for walking and cycling [153, 155, 156] building age, public transport provision, safety, crime and street connectivity [151] promoting alternate modes of transportation-public transportation, walking or biking - to increase activity levels [152, 158].

Access to healthy foods has been studied using GIS to identify ‘Food Deserts’ and measure the exposure to food outlets and retail stores to analyze the supply of food in a neighborhood [159-161] and to understand the environment influence on eating behaviors, as studied by French et al. [166], can facilitate the creation of interventions that change neighborhood environments in a way that promotes the intake of healthy food [167].

Health Technologies and Tools

GIS software packages of varying content and origin are widely available today. Some GIS software has been made for special purposes and limited analysis capabilities to that purpose, e.g. CMAP (Crime Mapping and Analysis Program) or LOLA (Locational...
analysis), whereas other more comprehensive packages provide a more general analysis tool, e.g. ArcGIS or MapInfo. In all GIS software on the commercial, or the free market, the core is the analysis they provide to the user. The GIS analysis enables the user to analyze why things are located where they are and how things are related. GIS analysis can improve data accuracy and create information that was previously unavailable and thereby help users get a deeper understanding of a place, make the best choices and prepare for future events and conditions.

GIS has a wide range of analysis options, which can be helpful for health prevention and promotion initiatives, disease curing and rehabilitation. These options include distance, neighborhood, smoothing, spatial interpolation, geostatistical (i.e. Geary’s C and Moran’s I), spatial relationship, density and geographic process analysis.

Examples of the impact of distance on health, e.g. Acharya et al. [91] among others [77, 78, 89, 164-167] where distance analyses is used to understand how utilization of services are influenced by distance to health services.

Variance of Geostatistical methods used to compare neighborhoods is often based on the diseased individual’s address, and relies on a process called geocoding which assigns latitude and longitude coordinates to the address. Different approaches to geocoding exist, with varying levels of accuracy and precision, impacting final analytic results. Zinszer et al. [168] among others [169, 170], evaluated the overall impact of accuracy and precision in geocoding [172-174] who also argue that further research into errors in health data geocoding is essential due to the uncertainty of their results. Improvements in geocoding was demonstrated in studies comparing geocoding methods delivered by the large providers [39, 175-177], which is what Lin et al. [178] did, using ESRI’s ArcGIS geocoder and Google maps’ geocoding service.

Another limitation of Geocoding patient’s residence ignor other locations where individuals spend significant amount of time: work, school, etc. Mapping people’s daily exposure is difficult without accurate knowledge of their behaviour and temporal locations; e.g. Whyatt et al. [179] used GPS to estimate people’s exposure to air pollution. Small sized GPS receivers are optimal for tracking movement and when combined with additional health-related sensors (e.g. ECG or heart rate monitors), can be powerful tools for measuring health-related parameters.

Apart from geocoding on individual levels it can be useful to geocode health survey data to small area levels, such as census tracts or even countries. Other data geocoded to the same geographic level can then be merged into the dataset for subsequent inferential analysis. A GIS can play a key role in merging these multiple data layers. As such uncertainty in geocoding and its potential implication in spatial analysis with health data should be considered in future studies.

A rising concern within health departments are the growing numbers of vulnerable population such as senior citizens, who would prefer to age in place. Aging in place refers to, “the ability to live in one’s own home and community safely, independently, and comfortably, regardless of age, income or ability level” [180]. One study looked at how combinations of GPS with other sensors were used to monitor a sample of elderly and disabled individuals remotely. An alert would be triggered and rescue personnel dispatched to the individual’s GPS location when certain criteria were met [176], enabling [181], these people to live more independently. Worringham et al. [182] combined single-lead ECG and heart rate sensors with GPS to help monitor patients in rehabilitation programs living who live in remotely areas or far from a hospital to participate from a distance and still have contact with qualified exercise scientists. The sensors provided the exercise scientists with information about the physiological parameters of a patient’s performance, while the GPS provided speed and location. Other technologies are also commonly used to measure physical activity such as accelerometers [152], pedometers [183] and mobile phones [184]. The exercise scientists can employ these tools to analyze the patient’s progress and give them advice in the ongoing rehabilitation process.

Other examples of remote sensing or field measurements via GPS. Aids in the Analyses of disease incidence rates and causes of infection impacted often rely on the exploration of environmental risk factors. Remote sensing, in general, refers to the collection of information about an object of phenomenon without making physical contact with it refers to the use of aerial sensor technologies to detect and classify objects on Earth (both on the surface, and in the atmosphere and oceans) by means of propagated signals (e.g. electromagnetic radiation). For example, Chunxiang et al. [190] analyzed the potential endemic disease risk after an earthquake. In practice, this information may be obtained from numerous kinds of monitors or from aerial imagery and satellite systems enabling disease monitoring, surveillance, control, evaluation and risk mapping. Allen and Wong [183] monitored and evaluated disease risk for West Nile virus using GIS and remote sensing to characterize and analyze the habitat and environmental conditions affecting mosquito breeding. Key data is able to be incorporated into the analysis which otherwise may not have been possible before: data at different spatial resolutions (e.g. vegetation types, land cover, surface temperatures, soil moisture and building density). , Johnson et al. [127] use remote sensing to obtain land surface temperatures and combined with sociodemographic risk factors to improve the understanding of health risks caused by high temperatures. And Khormi and Kumar [60] used remotely sensed data to understand housing density, street width and roof areas of houses. When this was combined with socioeconomic parameters, it represented a powerful explanation of risk for infectious disease [185]. The use of remote sensing to give an early warning of epidemic diseases within health research is relatively new, Herbreteau et al.’s [186] and Beck et al.’s [187]

**Discussion**

The health hazards that humans are exposed to every day are caused by a variety of aspects in their lives. Individuals are born with certain genes, making them more or less susceptible to certain diseases. In the course of an individual’s life his or her socioeconomic status, social conditions and relations affect the individual’s living conditions, which
again influence the individual’s lifestyle in terms of housing, mobility, nutrition, stimulants etc. Whether the health hazards are a result of one’s genes, childhood, socioeconomic status, social conditions or relations, living conditions or lifestyle, reducing the number of negative impacts is important and challenging. Such an effort involves a variety of sciences and subject matter experts who, through a collaborative team science approach, will help to achieve a fuller understanding of disease and health. Additionally, scientists need to be taught to think spatially, lest the use of GIS, remote sensing and GPS in health research remain limited [189, 190].

The geographical contribution to understanding disease occurrence, causality and spread rests in the relationship between disease distribution and environmental factors. Current technologies support more advanced calculations of environmental risk by including a more complete personal location history (e.g. historical address histories and workplace histories). Ultimately, by applying the use of GPS, genetic sequencing, and personal health sensors, it will be possible to conduct real-time exposure measurements.

For the past decade, climate changes have been an important topic, and some areas have witnessed an average rise in temperature, increased or decreased amounts of rain and more extreme weather phenomena as an indication of the climatic changes. These changes will most certainly affect the incidence and transmission of some diseases, and this can lead to further understanding of the way diseases are affected by climate and lead to more accurate placement of interventions. Vector-borne diseases, in particular, are climate sensitive. The ability to model climate changes can predict new endemic areas of risk and will inform future intervention focci. Along with the climate, human interference results in ongoing environmental changes, and especially in developing countries there is an increased risk of disease enhanced by development activities, demographic changes and the introduction of new products.

Prediction of disease outbreaks can be valuable information in reducing disease spread. Analysis for optimizing the location of health facilities and the utilization of these facilities are common in the research literature, but, as stated by Higgs [76], health outcomes constitute a relevant analysis factor in getting an estimate of the quality delivered by the health facilities, and not just a quantitative measurement. As many studies reveal, distance does have an impact on the accessibility of health facilities. Analysis of the impact of distances to health facilities on health outcomes as well as the efficiency of emergency services and the quality they deliver in relation to emergency response distance/time could provide valuable information for understanding human behaviour in the health care system. This in turn, could possibly lead to new performance indicators for allocating resources and facilities in health care planning. For example, how does accessibility affect utilization and health outcomes, when analysis is based on both public and private transportation? In several major European cities, public transportation is gaining ground as a priority for health promotion and ecological responsibility though reductions in air pollution and carbon footprint. Other aspects of city planning and “health in all policies” approaches to promote health include neighborhood factors that promote walking and running, the design of cities that are suitable for cycling, and noise and air pollution reduction in cities by restructuring traffic.

Temporal analysis tools and illustrations have become more common in GIS software packages and are therefore more easily accessible than previous; they can be applied to understand the effects of health interventions. Evaluating interventions to reduce or prevent disease incidence and prevalence can benefit from temporal analyses of the status of the disease before and after interventions. Analysis of pre-intervention disease status is an essential element in understanding the effect of interventions on disease occurrence both in the short and long term. It must also be noted that modern statistical software packages are getting more and more GIS functionalities, and GIS software packages are being supplemented with statistical functionalities mainly because spatial statistics often make use of different algorithms than traditional statistics applied in other sciences.

Health is not only about disease analysis; it is also about promoting aspects of a healthy lifestyle, such as healthy nutrition and regular physical activity. Often the promotion of healthy nutrition focuses only on making the healthier alternatives accessible, but individuals are still in a position where they can choose between the healthy and unhealthy alternatives. Today, the unhealthy alternatives are often presented as the easiest choice while the healthiest options are seldom placed in a way that makes the rational and healthier choice easy for the customers. Physical activity is proven to be healthy for the metabolism, but as people cannot be forced to exercise, other solutions that encourage physical activity are needed. Many initiatives without a geographic component have been developed to promote physical activity and geo-applications for mobile phones that encourage physical activity through the mapping and measuring of speeds and distances in training sessions are particularly prevalent.

The analysis of all the articles in the category “spatial analysis of disease” revealed that the majority focus on infectious diseases, which has a straightforward explanation as the spread of infections have a natural geographical component. Studies of malaria and schistosomiasis were the most frequently analyzed diseases. Malaria is also the most common cause of death and schistosomiasis – also known as bilharziosis – is a highly contagious disease that leaves the surviving victims with serious disabilities. It is transmitted through freshwater, which is vital to human life. Studies of dengue fever and HIV are represented equally with 25 and 22 studies, and Leishmaniasis was studied in 15 articles. It is noteworthy that four of the five most frequently studied diseases using GIS are spread by pests including three flying insects.

Of the non-infectious diseases, cancer is by far the most frequently studied, with more than twice as many articles than the second most studied, asthma with only 14 studies. It is significant that the studies of infectious diseases predominantly focus on the spread of the diseases and the causes are well known, whereas the studies on non-infectious diseases pursue the etiology of the disease by searching patterns in the prevalence and incidence rates in relation to a geographical component.

An explanation of this finding could be that infectious diseases are a result of an easily identifiable cause (cf. a mosquito),
which can be studied relatively easily. When studying non-infectious diseases the level of complexity increases significantly, and is more complicated to follow as they develop through a life span. Furthermore if effect sizes are small and distributed among variables it can contribute to a publication bias.

The search revealed very few studies on preventive activities and health outcomes via GIS. This might be due to the fact that preventive studies are much different than other health related studies. In general there are much fewer studies on prevention - those studies take a long time if prospective and require a lot of variables if retrospective where geography is just one. In addition outcomes can be difficult and often surrogates are used for the outcomes of interest.

Obtaining accurate and reliable health data can be a lengthy and expensive process. One that is often confounded by privacy and confidentiality requirements which need to be balanced with the need for good quality analysis. Further, exclusion of key data sets due to privacy concerns or data being withheld due to commercial or political issues skews research results. In different countries there are different traditions in the registration of patients, which entails a variety of barriers in geocoding patient records. North European countries have a tradition for registering a large amount of information for every citizen, which potentially allows for easier process in geocoding patient registers if location is included in this plethora of information. In other parts of the world financing issues tends to restrict the availability of accurate databases and the geocoding process will require a lot of time and resources. Some countries have legally limited the data that each individual is permitted to register at a government level, which means that geographical references may not be recorded in patient records in some jurisdictions. Regardless of the problem each country has in geocoding their patient records, consideration for patient confidentiality is an issue when analyzing the data and making results publicly available. This requires the data providers to ensure the data are appropriately aggregated either through legislation or pre-aggregation of the data.

It is well known that health problems are related to a number of social environmental variables such as poverty and other contextual variables, which in many cases are available at some scales of analysis. However we find a great variety in the granularity of the available data. In Scandinavia data are readily available on an individual level, and simple rules for anonymization and ethics are regulating research on these data. In many other countries only very aggregate data are available, and the use of the data can be complicated to get access to due to strict legislation. An elaboration of this discussion would require further studies of various health surveillance data and their sources to point out limitations and benefits.

Despite the advancement of the sophisticated tools for analyzing GIS data, reliable and valid health data will still have to be provided. The primary use of health data is in the clinic where data are recorded and stored for the purpose of delivering health services to the individual. Applying health data in GIS analysis can be considered as secondary use of health data, which raises a number of concerns. Charles Safran et al. reported the outcome of a panel of diverse stakeholders and experts discussing a range of issues related to secondary use of health data [191]. Using health data outside direct health care delivery raises a number of complex ethical, political, technical and social issues. The panel recommended a number of initiatives that could also apply to GIS analysis of health data. Ultimately, the responsibility for ensuring privacy and safeguarding patient data applies to all data users, and when data are attached to a specific location this responsibility is significant. The de-identification and anonymization of data should be addressed in all GIS applications.

After the 2011 limit of the search in this paper there has been two papers reviewing mapping and surveillance of infectious disease [192, 193]. As they have a slightly different aim, they do not give rise to revision of the results and discussion in this paper.

Conclusion

The increasing number of publications related to GIS and health demonstrates the major growth this research area has seen recently. As our proposed framework illustrates, a lot of research has been undertaken to understand disease and health, but as our understanding increases, new questions arise along with the development of new technologies and applications thereof, which further increase the possibilities for measuring, analysing and understanding disease and health. This framework does not specifically cover studies of realized benefits of applying GIS analysis to health related data, hence future studies would be required to investigate outcomes, utility and benefits. Furthermore it might be relevant to perform further in depth studies on specific aspects of which diseases are surveyed on different continents. The framework has found and systematized a large number of recent articles concerning GIS and health, and the complete set of references, is available as supplementary material here: http://www.schattauer.de/fileadmin/assets/zeitschriften/IMIA_Yearbook/A_complete_list_of_literature_found_for_the_article.pdf.

References

1. Hipocrates, Clifton F, Watts J. Hippocrates upon air, water, and situation: upon epidemic diseases; and upon prognosticks, in acute cases especially. To this is added (by way of comparison) Thucydides’s account of the plague of Athens, the whole translated, methodized, and illustrated with useful and explanatory notes. London; Printed for J. Watts; 1734, 389 p.
A Review and Framework for Categorizing Current Research and Development in Health Related Geographical Information Systems (GIS) Studies


A Review and Framework for Categorizing Current Research and Development in Health Related Geographical Information Systems (GIS) Studies


142. Kligerman M, Sallis JF, Ryan S, Frank LD, Nader PR. Association of neighborhood design and recreation environment variables with physical activity and body mass index in adolescents. Am

IMIA Yearbook of Medical Informatics 2014

Correspondence to: Anders Knør Lyseen Department of Development and Planning Aalborg University Aalborg, Denmark E-mail: alyseen@plan.aau.dk

IMIA Yearbook of Medical Informatics 2014